

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO  
FAKULTETA ZA MATEMATIKO IN FIZIKO

Marko Verdenik

**Ločevanje skupin elektromiografskih  
posnetkov maternice s terminskim  
in prezgodnjim porodom z uporabo  
koherenčne funkcije**

DIPLOMSKO DELO  
NA INTERDISCIPLINARNEM UNIVERZITETNEM  
ŠTUDIJU

MENTOR: prof. dr. Franc Jager

Ljubljana 2015



Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljane ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

*Besedilo je oblikovano z urejevalnikom besedil  $\LaTeX$ .*



Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Tematika naloge:

Razvijte postopke klasifikacije skupin posnetkov elektromiograma maternice s terminskim in prezgodnjim porodom z uporabo koherenčne funkcije. Za oceno zmogljivosti ločevanja skupin posnetkov snemanih zgodaj (23. teden) in pozno (32. teden) v nosečnosti uporabite mednarodno referenčno podatkovno bazo elektromiogramov maternice TPEHG DB. Uporabnost koherenčne funkcije za ločevanje skupin posnetkov karakterizirajte z uporabo analize variance, ANOVA. Posnetke klasificirajte na podlagi mediane amplitude in integrala koherenčne funkcije močnostnih spektrov v različnih frekvenčnih področjih. Ocenite zmogljivost klasifikacije z uporabo Bayesovega klasifikatorja, odločitvenih dreves in lastno razvitega klasifikatorja.



## IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Marko Verdenik, z vpisno številko **63070471**, sem avtor diplomskega dela z naslovom:

*Ločevanje skupin elektromiografskih posnetkov maternice s terminskim in prezgodnjim porodom z uporabo koherentne funkcije*

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom prof. dr. Franca Jagra,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela,
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, 5. marca 2015

Podpis avtorja:

*Prof. Francu Jagru se zahvaljujem za ponujene izzive, čas in usmerjanje ob nastajanju diplomske naloge. Očetu Ivanu za nasvete in spodbudne besede in mami Darki za potrpežljivost.*



Imagination is more important  
than knowledge.

For knowledge is limited to all  
we now know and understand,  
while imagination embraces  
the entire world, and all there  
ever will be to know and  
understand.

---

Albert Einstein

# Kazalo

Seznam uporabljenih kratic

Povzetek

Abstract

<b>1</b>	<b>Uvod</b>	<b>1</b>
1.1	Namen dela . . . . .	4
<b>2</b>	<b>Metode</b>	<b>5</b>
2.1	Elektromiografija (EMG) . . . . .	5
2.2	Podatkovna baza TPEHG DB . . . . .	6
2.3	Predobdelava signalov . . . . .	9
2.3.1	Dvosmerna faza filtriranja . . . . .	9
2.3.2	Butterworthov filter . . . . .	10
2.3.3	Pasovna območja filtriranja signalov . . . . .	11
2.4	Koherenčna funkcija . . . . .	12
2.4.1	Welcheva metoda . . . . .	14
2.4.2	Vrednotenje koherenčne funkcije . . . . .	15
2.5	Analiza variance ANOVA . . . . .	17
2.6	Klasifikacija posnetkov . . . . .	19
2.6.1	Klasifikator z uporabo značilk v različnih frekvenčnih območjih . . . . .	20
2.6.2	Naivni Bayesov klasifikator . . . . .	24
2.6.3	Odločitvena drevesa . . . . .	24

## KAZALO

2.6.4	Ocena delovanja klasifikatorja . . . . .	25
<b>3</b>	<b>Rezultati</b>	<b>29</b>
3.1	Ločevanje skupin . . . . .	30
3.1.1	Ločevanje zgodaj snemanih posnetkov z uporabo mediane amplitude . . . . .	31
3.1.2	Ločevanje zgodaj snemanih posnetkov z uporabo integrala . . . . .	32
3.1.3	Ločevanje pozno snemanih posnetkov z uporabo mediane amplitude . . . . .	33
3.1.4	Ločevanje pozno snemanih posnetkov z uporabo in- tegrala . . . . .	35
3.2	Ocenjevanje klasifikacije . . . . .	36
3.2.1	Ocena zmogljivosti delovanja klasifikatorja z uporabo značilk v različnih frekvenčnih območjih . . . . .	37
3.2.2	Ocene zmogljivosti delovanja Bayesovega klasifika- torja . . . . .	38
3.2.3	Ocene zmogljivosti delovanja odločitvenega drevesa	41
<b>4</b>	<b>Zaključek</b>	<b>44</b>
<b>5</b>	<b>Dodatek</b>	<b>47</b>
A	Statistični rezultati ANOVE . . . . .	47

## Seznam uporabljenih kratic

KRATICA	angleško	slovensko
ANOVA	analysis of variance	analiza variance
EMG	electromyography	elektromiografija
EHG	electrohysterography	elektrohisterografija
TPEHG DB	term-preterm	podatkovna baza terminskih in
	electrohysterogram database	prezgodnjih elektrohisterogramov
PE	preterm early	prezgodnji snemano zgodaj
TE	term early	terminski snemano zgodaj
PL	preterm late	prezgodnji snemano pozno
TL	preterm early	terminski snemano pozno
FIR	finite impulse response	končni odziv na enotin impulz
IIR	infinite impulse response	neskončni odziv na enotin impulz
TP	true positive	resnično pozitiven
FP	false positive	lažno negativen
TN	true negative	resnično negativen
FN	false negative	lažno negativen
Se	sensitivity	občutljivost
Sp	specificity	specifičnost
Acc	accuracy	natančnost

# Povzetek

V diplomskem delu predstavimo avtomatično analizo elektromiograma maternice s koherenčno funkcijo, ki je ena izmed nelinearnih tehnik procesiranja signalov. Uporabili smo posnetke mednarodne referenčne baze TPEHG DB, ki vsebuje 300 posnetkov. Signale smo predprocesirali z devetimi različnimi Butterworthovimi pasovno-prepustnimi filtri in v izogib faznemu popačenju uporabili dvosmerno shemo filtriranja.

Ločevanje skupin je potekalo v dveh variantah, ločevanje med zgodaj snemanimi in ločevanje med pozno snemanimi posnetki. Izračunali smo koherenčno funkcijo med vsemi pari posnetkov za vsako od variant. Računali smo jo med močnostnima spektroma signalov. Za oceno koherence za celotno frekvenčno območje smo izbrali dve cenilki - mediano amplitude in integral.

Enosmerna analiza varianca ali ANOVA je pokazala, katere skupine posnetkov so primerne za ločevanje prezgodnjega in terminskega poroda. Frekvenčna območja in signale, katerih p-vrednosti so manjše od 0,05, smo uporabili za klasifikacijo posnetkov. Za klasifikacijo smo uporabili Bayesov klasifikator, odločitvena drevesa in klasifikator, ki smo ga empirično sestavili sami. Opazili smo, da se koherenca med terminskimi porodi med frekvenčnima območjema 1-2,5 Hz in 2,5-3,5 Hz znatno zmanjša, medtem ko se koherenca prezgodnjih porodov znatno ne spremeni. Ocenjevanje zmogljivosti klasifikacije je potekalo na tri načine - na učni množici, po principu učna-testna množica in s pristopom "izpusti enega". Najboljšo oceno klasifikacije smo dobili z uporabo odločitvenih dreves na učni množici, na frekvenčnem območju 0,3-2,5 Hz, kjer je bila

občutljivost 95 %, specifičnost in natančnost pa 98 %. Malo slabše rezultate smo dobili z uporabo lastnega klasifikatorja. Občutljivost je bila med 58 % in 63 %, specifičnost pa med 61 % in 65 % za izbrane filtre in kanale. Klasifikacija z Baysovim klasifikatorjem pa ni pokazala vzpodbudnih rezultatov, z občutljivostjo blizu 0 %.

**Ključne besede:** prezgodnji porod, elektrohistogram, koherenčna funkcija, ločevanje skupin, klasifikacija, klasifikator

# Abstract

In this thesis we present automatic analysis of electromyogram of uterus (electrohysterogram) using coherence function which is one of non-linear signal processing techniques. We used records of international reference database TPEHG DB (Term-Preterm Electrohysterogram DataBase), which contains 300 electrohysterogram records. We preprocessed signals with nine different band-pass Butterworths filters with forward-backward filtering to avoid zero phase shift.

Separation of groups took place in two variants, among early recorded and among late recorded records. We calculated coherence function between all pairs of records for each of variants. For calculation we used power spectrum of signals. Coherence estimation for whole frequency range, was made with two techniques - median amplitude and integral.

Analysis of variance or ANOVA showed which frequency ranges and signals are useful for preterm - term records separation. For records classification we used frequency intervals and signals with p-value less than 0,05. Evaluation of classification was made on Bayes classifier, decision trees and our own built classifier. We developed it empirically, based on coherence decreasing among term records from frequency range 1-2,5 Hz to 2,5-3,5 Hz. Performance evaluation of classification is done in three ways - on training set, on the principle of training-testing set and with the approach "omitted one". Best results were shown with decision tree at frequency range 0,3-2,5 Hz, where sensitivity was 95 %, specificity and accuracy were 98 %. With our own developed classifier we reach sensitivity between 58 % and 63 % and specificity between 58 %

and 63 %. Classification using Bayes classifier did not show good results, having sensitivity close to 0 %.

**Key words:** preterm delivery, electrohysterogram, coherence function, group separation, classification, classifier



# Uvod

Prezgodnji porod pri človeku je porod otroka pred 37. tednom nosečnosti. V Sloveniji se je med leti 2008 in 2012 kar 6,2 % novorojenčkov rodilo prezgodaj [1]. Nedonošenčki so podvrženi večjemu tveganju za kratkoročne in dolgoročne zaplete, ki lahko prizadenejo dihala, prebavila, sluh, vid, živčni in imunski sistem [11]. V ekstremnih primerih lahko pride do cerebralne paralize ali duševne zaostalosti. Potrebna je dodatna nega, da se jim omogoči čim bolj normalno življenje, cena katere je lahko zelo draga. Včasih celo preseže zaslužek nedonošenčka v celotnem življenju [4]. Prezgodnji porod težko preprečimo, lahko pa nosečnost podaljšamo s pomočjo medikamentov in tako plodu omogočimo zdrav razvoj. Bolj zgodaj odkrijemo, da bo do njega prišlo, več možnosti imamo za preprečevanje zapletov. Tako je zgodnje odkrivanje prezgodnjega poroda ključno za zmanjševanje števila nedonošenčkov. Torej manjše število oseb, ki bi potrebovale zdravstveno oskrbo in s tem velik finančni prihranek.

Dosedanje raziskave napovedovanja prezgodnjega poroda so pokazale, da obstajajo indici, ki so povezani s prezgodnjim porodom. Napovedne modele delimo v dve skupini. V prvo spadajo modeli, ki temeljijo na opazovanju vidnih dejavnikov, druga skupina napovednih modelov pa temelji na analizi elektromiograma maternice. Vidni dejavniki, ki povečujejo tveganje za prezgodnji porod, so: kajenje, povišan pritisk, sladkorna bolezen, prirojene nepravilnosti maternice, kratek maternični vrat in drugi [11, 8].

Z elektromiogramom maternice merimo njeno električno aktivnost. Mehanska aktivnost maternice, ki jo sprožajo spontani krči, je povezana

z električno aktivnostjo. Med samim porodom povečana električna aktivnost v EMG signalu maternice sovpada s povečanjem pritiska, ki je znak za popadek. Presenetljivo pa je električna aktivnost maternice, ki je povezana s popadki, prisotna že zgodaj v nosečnosti. Spontane popadke so identificirali že v 19 tednu nosečnosti [12]. Nivo električne aktivnosti se poveča, ko se približuje čas poroda, še bolj izrazito pa 3 do 4 dni pred porodom [13]. Prav značilnosti krčev, čas začetka, pogostost in intenziteta krčev, se pri tistih nosečnicah, kjer je tveganje za prezgodnji porod večje, razlikujejo od krčev nosečnic, ki rodijo ob roku. Značilnost krčev se na elektromiogramu kažejo v različnih frekvencah in amplitudah električne aktivnosti. Zajem posnetkov EMG maternice je neinvazivna metoda in bi jo bilo preprosto uresničiti v klinični praksi.

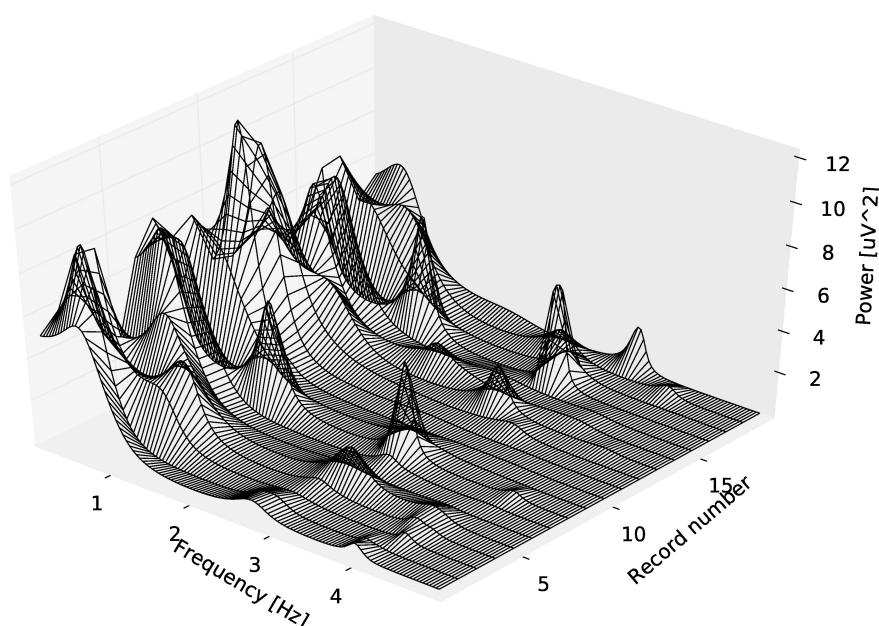
Uporaba analize EMG [2] za ločevanje prezgodnjih in terminskih porodov temelji na različnih linearnih in nelinearnih tehnikah za procesiranje signalov. Linearne tehnike so koren srednje vrednosti, vrh in mediana frekvenca močnostnega spektra signala. Nelinearne tehnike pa so maksimalni Lyapunov eksponent, korelacija in vzorčna entropija. Analiza je bila opravljena na isti bazi posnetkov kot naša diplomska naloga, brez dodatnih frekvenčnih območij. Tehniki, ki sta se izkazali kot najboljši, sta mediana frekvenca in vzorčna entropija pri uporabi pasovno prepustnega filtra 0,3-3 Hz.

Tehnika	Sig.	PE - TE			PL - TL		
		1	2	3	1	2	3
Kvadrat sr. vrednosti		0,586	0,361	0,636	0,349	0,141	0,612
Frekvenčni vrh		0,630	0,252	0,138	0,100	0,201	0,176
Mediana frekvence		0,371	0,696	<b>0,030</b>	0,059	0,568	0,212
Avtokorelacija		0,085	0,089	0,327	0,897	0,340	0,614
Lyapunov koeficient		0,543	0,533	0,670	0,518	0,175	0,743
Korelacija		0,150	0,676	0,790	0,961	0,377	0,976
Vzorčna entropija		0,326	0,882	<b>0,035</b>	0,172	0,184	0,165

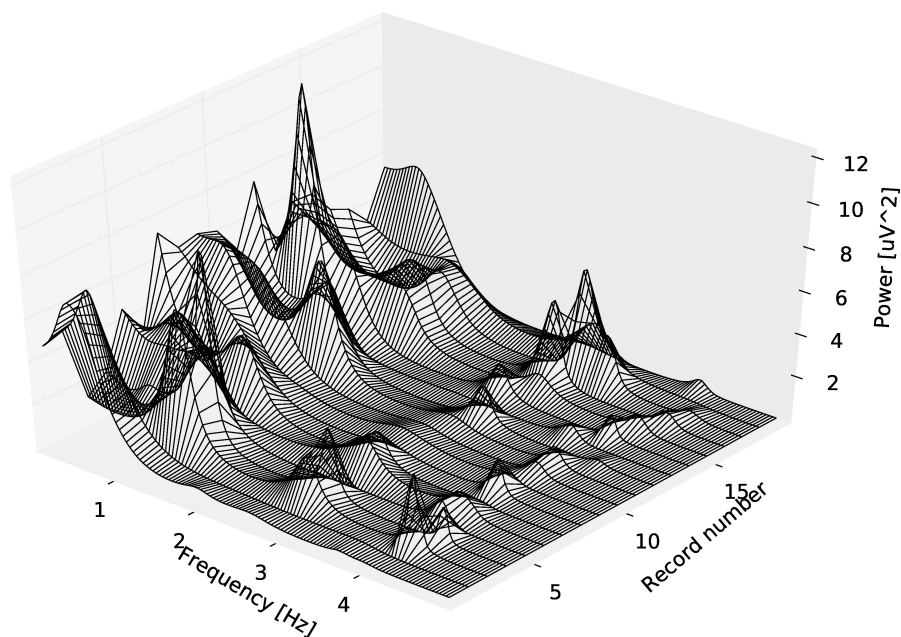
Tabela 1.1: Vrednotenje (p-vrednosti) tehnik za ločevanje skupin [2]

Koherenčna funkcija kot nelinearna tehnika za obdelavo signalov je dala vzpodbudne rezultate na drugih fizioloških signalih [10]. Njena značilnost iskanja skupnih frekvenc med dvema signaloma in ocenjevanje njune podobnosti nakazuje, da bi bila primerna tehnika za ločevanje med prezgodnjimi in terminskimi porodi. Osnovna ideja je bila, da si bodo posnetki, kjer je prišlo do prezgodnjega poroda, frekvenčno bolj podobni med seboj, kot s posnetki iz skupine, kjer je bil porod ob roku. Nelinearne tehnike so računsko zahtevnejše od linearnih, vendar to ne igra velike vloge pri napovedovanju prezgodnjega poroda, ker ga ni potrebno izvajati v realnem času.

Salve električne aktivnosti maternice, ki sovpadajo s popadki, imajo amplitudni vrh med 0,34 Hz in 1 Hz [9]. Na slikah 1.1 in 1.2, močnostnih spektrov EMG 19 nosečnic vidimo, da so aktivna tudi druga frekvenčna območja. Da bi videli ali ta frekvenčna območja vsebujejo informacije za napoved prezgodnjega poroda, smo posebej opazovali tudi višja frekvenčna območja signalov.



Slika 1.1: Močnostni spekter 19 posnetkov EHG, ki so se končali s prezgodnjim porodom [15]



Slika 1.2: Močnostni spekter 19 posnetkov EHG, ki so se končali s porodom ob roku [15]

## 1.1 Namen dela

Naši cilji so bili naslednji:

- Raziskati uporabnost koherenčne funkcije za ločevanje skupin posnetkov EMG maternice s prezgodnjim in terminskim.
- Pokazati, da so za ločevanje skupin, uporabna tudi višja frekvenčna območja;
- Klasificirati posnetke v eno izmed skupin na podlagi koherenčne funkcije z uporabo Bayesovega klasifikatorja in odločitvenih dreves.
- Razvoj preprostega lastnega klasifikatorja.

# Metode

## 2.1 Elektromiografija (EMG)

EMG je tehnika za določanje električne aktivnosti mišic s pomočjo elektromiografa, ki meri električni potencial mišičnih celic ob krčenju mišic. Osnovni način merjenja je z uporabo elektrod. Ob krčenju mišic se ustvari električni potencial, ki ga elektrode zaznajo in signal preko ojačevalnika pošljejo in zapišejo v računalnik. Posnetek signala se imenuje elektromiogram.

Beseda *myos* v grščini pomeni mišica, med katere spada maternica, ki se v grščini imenuje *hystera*. Zato pri določanju električne aktivnosti maternice namesto elektromiografije uporabljamo bolj specifičen izraz elektrohisterografija in podobno za posnetek signala elektrohisterogram.

Začetki elektromiografije segajo že v leto 1666 z zapisi Francesca Redija. Objavil je, da posebna vrsta mišice električnega skata lahko generira elektriko. Leta 1771 Luigi Galvani v svojem delu *De Viribus Electricitatis in Motu Musculari Commentarius* pokaže, da električna stimulacija mišičnega tkiva povzroča mišično kontrakcijo. Kljub temu pa električni odziv mišice ni bil zabeležen vse do razvoja napredne snemalne opreme. Prvi moderni EMG je bil izdelan leta 1942 na McGillovi univerzi v Montrealu.

Poznamo invazivno in neinvazivno metodo. Pri invazivni se v mišico zabode ena ali več elektrod, ki zaznavajo prisotnost, velikost in obliko signala. Pri tem načinu zaznavamo signale le v posameznih mišičnih

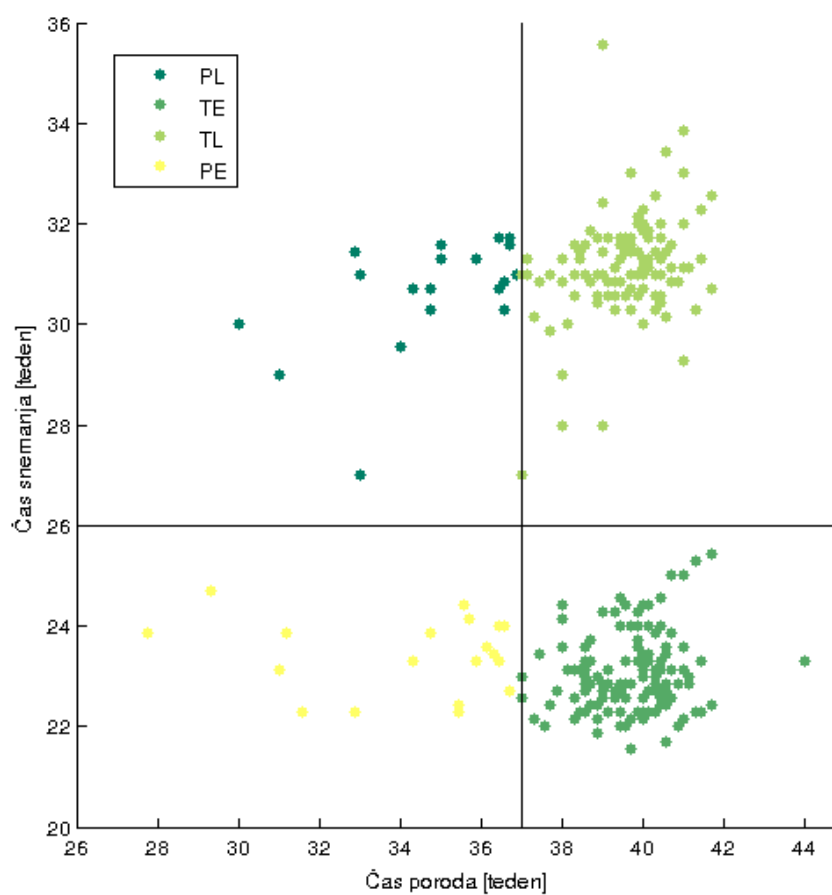
vlaknih. Ker je metoda invazivna, je boleča in lahko pride do lažjih poškodb. Pri neinvazivni metodi se na kožo nad mišico pritrdi 2 elektrodi, ki nato merita električno aktivnost celotne mišice.

## 2.2 Podatkovna baza TPEHG DB

Podatkovna baza TPEHG DB [2] (ang. Term-Preterm ElectroHystero-Gram DataBase) vsebuje posnetke elektrohistrogramov (elektromiogram maternice), ki so bili posneti na Univerzitetnem kliničnem centru Ljubljana na oddelku za klinično perinatologijo med leti 1997 in 2005. Posnetki so bili zajeti ob redni kontroli nosečnice, okoli 22. ali okoli 32. tedna nosečnosti.

Podatkovna baza TPEHG vsebuje 300 EMG posnetkov maternice. En posnetek pripada eni nosečnici. Posnetki se delijo v naslednje 4 skupine:

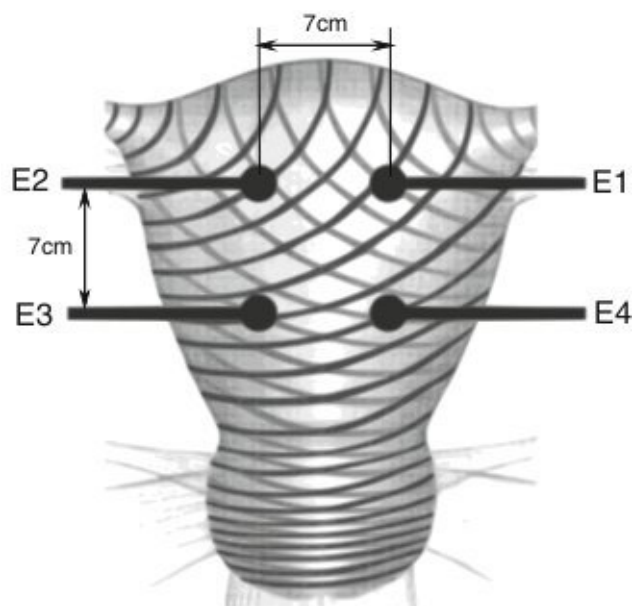
- prezgodnji porod snemano zgodaj (PE - Preterm Early): posnetki zajeti okoli 22. tedna nosečnosti, ki se je končala prezgodaj oz. pred 37. tednom nosečnosti;
- prezgodnji porod snemano pozno (PL - Preterm Late): posnetki zajeti okoli 32. tedna nosečnosti, ki se je končala prezgodaj oz. pred 37 tednom nosečnosti;
- terminski porod snemano zgodaj (TE - Term Early): posnetki zajeti okoli 22. tedna nosečnosti, ki se je končala v roku oz. po 37. tednu nosečnosti;
- terminski porod snemano pozno (TL - Term Late): posnetki zajeti okoli 32. tedna nosečnosti, ki se je končala prezgodaj oz. po 37. tednu nosečnosti.



Slika 2.1: Časovna razporeditev posnetkov

Posnetki so bili pridobljeni s pomočjo 4 elektrod. Elektrode so bile nameščene po sledeči shemi:

- prva elektroda (E1) je bila nameščena 3.5 cm levo zgoraj od popka,
- druga elektroda (E2) je bila nameščena 3.5 cm desno zgoraj od popka,
- prva elektroda (E3) je bila nameščena 3.5 cm levo spodaj od popka,
- druga elektroda (E4) je bila nameščena 3.5 cm desno spodaj od popka.



Slika 2.2: Postavitev elektrod na trebuhu [2]

Vsak posnetek je sestavljen iz treh signalov, ki jih dobimo z razliko električnega potenciala med elektrodami.

- Signal 1:  $E2 - E1$
- Signal 2:  $E2 - E3$
- Signal 3:  $E4 - E3$

Vsi posnetki so bili predhodno obdelani z nizkoprepustnim analognim filtrom z mejno frekvenco 5 Hz. Posamezen posnetek ima dolžino 30 minut. Frekvenca vzorčenja poseameznega signala na kanal je 20 vzorcev na sekundo. Merilno območje napetosti pa je v območju  $\pm 2.5 \text{ mV}$ .

Poleg originalnega signala je bil signal filtriran še s tremi različnimi digitalnimi filtri. Uporabljen je bil Butterworthov filter 4. reda z naslednjimi pasovno prepustnimi frekvencami:

- od 0.08 Hz do 4 Hz
- od 0.3 Hz do 3 Hz
- od 0.3 Hz do 4 Hz



## 2.3 Predobdelava signalov

Signali posneti z EMG maternice vsebujejo šum. EMG poleg električne aktivnosti maternice zajame še električne aktivnosti preostalih delov telesa, kot so mišice in srce. Da bi dobili izključno električno aktivnost maternice, je potrebno signal filtrirati, tako da se električne aktivnosti, ki ne pripadajo maternici, izločijo. Obstajata dve skupini filtrov:

- FIR - filter (filter s končnim odzivom na enotin impulz),
- IIR - filter (filter z neskončnim odzivom na enotin impulz).

Filtri z neskončnim odzivom na enotin impulz so računsko manj zahtevni pri podobni kvaliteti, vendar je z njimi nemogoče doseči linearni fazni zamik. Za ta problem obstaja rešitev dvosmerne filtriranja [16]. Signal filtriramo v eni smeri, ga obrnemo, filtriramo v drugi smeri ter nato obrnemo nazaj. To si lahko privoščimo samo, če ne potrebujemo obdelave v realnem času.

### 2.3.1 Dvosmerna faza filtriranja

$$z_1(n) = h(n) * x(n) \quad (2.1)$$

$$z_2(n) = h(n) * z_1(-n)$$

$$s(n) = z_2(-n)$$

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \quad (2.2)$$

$$x(-n) \xrightarrow{F} X * (e^{j\omega}) \quad (2.3)$$

$$\begin{aligned} S(e^{j\omega}) &= Z_2^*(e^{j\omega}) = H^*(e^{j\omega})Z_1(e^{j\omega}) \\ &= H^*(e^{j\omega})H(e^{j\omega})X(e^{j\omega}) \\ &= |H(e^{j\omega})|^2 X(e^{j\omega}) \end{aligned} \quad (2.4)$$

### 2.3.2 Butterworthov filter

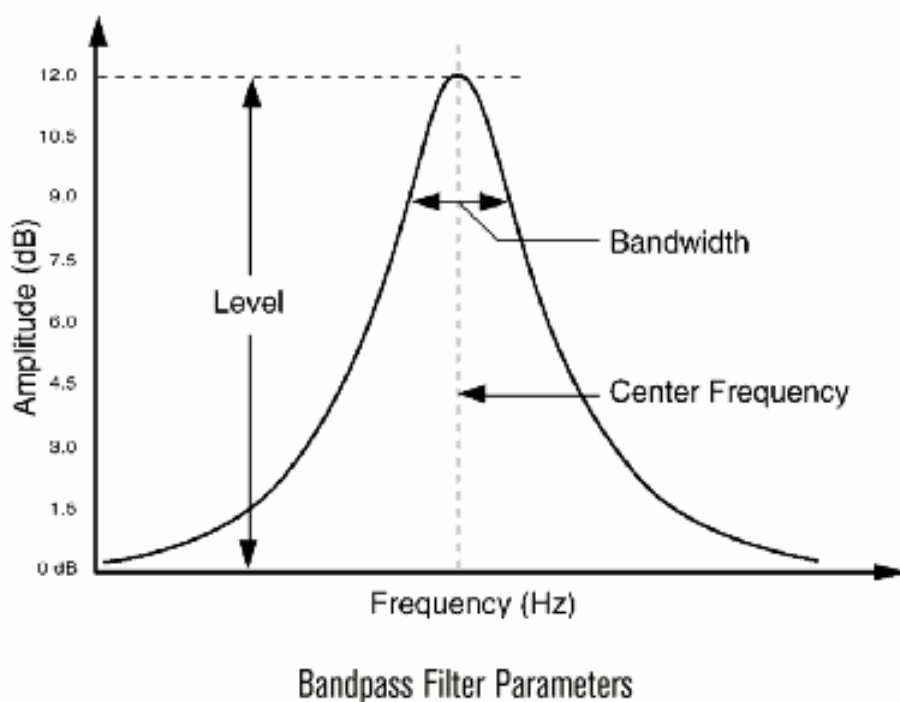
Je filter z neskončnim odzivom na enotin impulz. Prenosna funkcija nizko prepustnega Butterworthovega filtra je 2.5 in visoko prepustnega 2.6:

$$|H(\Omega)|^2 = \frac{1}{1 + \left(\frac{\Omega}{\Omega_C}\right)^{2n}} \quad (2.5)$$

$$|H(\Omega)|^2 = \frac{1}{1 + \left(\frac{\Omega_C}{\Omega}\right)^{2n}} \quad (2.6)$$

Njegova karakteristična funkcija v digitalni domeni je:

$$H(z) = \frac{B(z)}{A(z)} = \frac{b(1) + b(2)z^{-1} + \dots + b(n+1)z^{-n}}{1 + a(2)z^{-1} + \dots + a(n+1)z^{-n}} \quad (2.7)$$



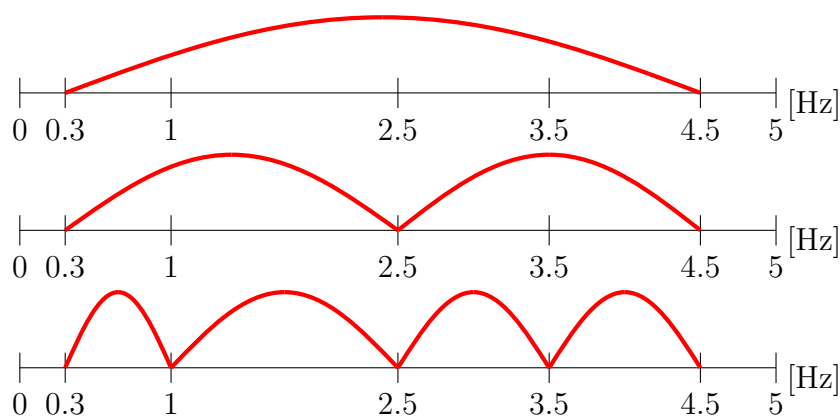
Slika 2.3: Frekvenčni odziv Butterworthovega filtra

Za implementacijo filtra smo uporabili, že obstoječe knjižnice programskega jezika MATLAB. V knjižnici za procesiranje signalov smo z

ukazom `[a,b] = butter(order,[w1 w2])` kreirali Butterworthov filter. Order je red filtra. Uporabili smo filter 4. reda,  $w_1$  je spodnja mejna frekvenca in  $w_2$  zgornja mejna frekvenca.

### 2.3.3 Pasovna območja filtriranja signalov

Dosedanje raziskave nelinearnih tehnik procesiranja signalov maternice so potekale na signalih iz baze, ki so bili filtrirani z izbranimi pasovno prepustnimi filtri. Tako so bila raziskana frekvenčna območja, za katere velja, da so najbolj pokazljiva pri problematiki napovedovanja prezgodnjega poroda. Pri svojem delu pa smo poleg že raziskanih frekvenčnih območij raziskali še ostala višja frekvenčna območja. Posnetki v bazi so bili filtrirani z analognim filtrom 5 Hz, zato smo se morali omejiti na frekvenčni pas do 5 Hz. Frekvence blizu 0 Hz in 5 Hz vsebujejo motnje, zato smo frekvenčno območje dodatno omejili na spodnjo mejo 0,3 Hz in zgornjo mejo 4,5 Hz. To območje smo hierarhično razdelili na sedem frekvenčnih območij, kot prikazuje slika 2.4.



Slika 2.4: Frekvenčna območja filtriranja signalov

Za frekvenčno območje od 0,3-4,5 Hz smo uporabili že obstoječ filtriran signal iz baze. Signal iz baze je bil filtriran z mejnima frekvencama 0,3 Hz in 4 Hz, ki pa je vseeno bil skladen z našim izbranim območjem. Uporabili smo tudi preostala dva obstoječa filtrirana signala iz baze, ker

so se pokazali dobri rezultati pri drugih nelinearnih tehnikah ločevanja skupin in zaradi preproste uporabe.

## 2.4 Koherenčna funkcija

Spektralna koherenca je statistika, ki se uporablja za preučevanje relacije med dvema signaloma ali med dvema podatkovnima množicama. Beseda koherenca izhaja iz latinske besede *cohaerentia* - naravna ali logična povezava. Koherenčna funkcija nam omogoča poiskati skupne frekvence in ovrednotiti podobnost signalov, ne da pa nam nobene informacije o času. Zavzame lahko vrednosti med 0 in 1. Vrednost 0 pomeni, da se signala popolnoma razlikujeta, vrednost 1 pa da sta si signala enaka.

Koherenčna funkcija ali koherenca je definirana kot:

$$C_{xy}(\omega) = \frac{|P_{xy}(\omega)|^2}{P_{xx}(\omega)P_{yy}(\omega)}, \quad (2.8)$$

kjer sta  $P_{xx}$  in  $P_{yy}$  močnostna spektra signalov  $x$  in  $y$ ,  $P_{xy}$  navzkrižni močnostni spekter za ta signala in  $\omega$  je frekvenca.

Močnostni spekter ali periodogram in navzkrižni spekter sta definirana kot

$$P_{xx}(\omega) = |\hat{x}(\omega)|^2 = \hat{x}(\omega)\overline{\hat{x}(\omega)} \quad (2.9)$$

$$P_{xy}(\omega) = \hat{x}(\omega)\overline{\hat{y}(\omega)}, \quad (2.10)$$

kjer je  $\hat{x}$  konjugirana vrednost  $x$  in

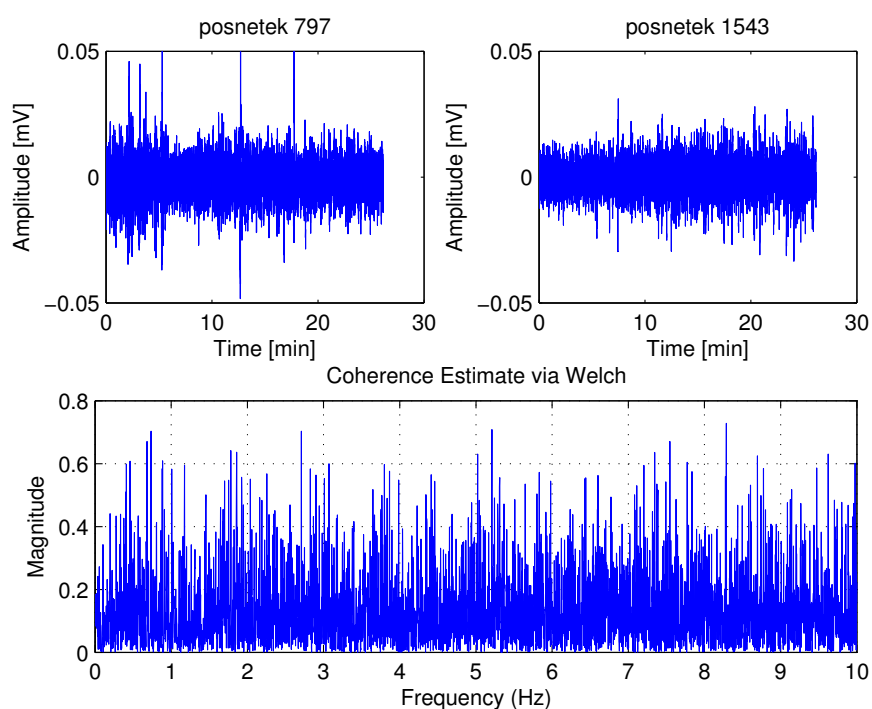
$$\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt \quad (2.11)$$

je Fourierova transformacija.

Poznamo dve metodi za izračun koherenčne funkcije: Welchova in MVDR (minimum variance distortionless response) metoda. Pri izračunavanju rezultatov smo uporabili že vgrajeno MATLAB funkcijo, ki uporablja Welschevo metodo.

`[Cxy,F] = mscohere(x,y>window,noverlap,nfft,fs)`

Parameter	Opis parametra	Privzeta vrednost
<b>nfft</b>	Dolžina FFT, ki določa frekvence, pri katerih je koherenca določena.	Največ 256 ali naslednja potenca števila 2, ki je večja od dolžine odseka.
<b>fs</b>	Frekvenca vzorčenja.	1
<b>window</b>	Velikost in vrsta okna.	Hammingovo okno dolžine, ki zajame 8 enakih odsekov x in y.
<b>noverlap</b>	Število vzorcev, ki se prekrivajo pri enem odseku.	Vrednost, ki zajame 50 % prekrivanja.



Slika 2.5: Primer koherenčne funkcije med dvema signaloma

### 2.4.1 Welcheva metoda

Vrh amplitude v navzkrižnem spektru pomeni, da je na tem mestu skupna frekvenca, ki je prisotna v obeh signalih. Kljub temu pa je mogoče, da se ta frekvenca pojavi v različnem času pri danima signaloma. Zatorej poleg navzkrižnega močnostnega spektra računamo koherenco na podlagi povprečne ocene močnostnega spektra posameznih signalov in navzkrižnega močnostnega spektra obeh signalov. Na ta način pridobimo informacijo o skupnih frekvencah ob istem času. To metodo je prvič predstavil Welsch leta 1967 [7].

Za diskretne signale (časovne vrste) izračunamo močnostni spekter za koherentno funkcijo na naslednji način. Signal razdelimo na  $K$  fragmentov dolžine  $M$  (možno prekrivanje)

$$x^l(n) = x(n + lD); \quad 0 \leq l \leq K - 1, 0 \leq n \leq M - 1 \quad (2.12)$$

in  $D$  je izbran zamik. Izbrani fragmenti signala so pomnoženi s funkcijo izbranega okna  $w(n)$  in za vse izračunamo močnostni spekter

$$P_{xx}^i(f) = \frac{1}{E_w} \left| \sum_{n=0}^{M-1} x^i(n) w(n) e^{-2i\pi(f/f_{pr})n} \right|, \quad (2.13)$$

kjer

$$E_w = \sum_{n=0}^{M-1} w^2(n) \quad (2.14)$$

Okno, ki smo ga uporabili, je Hammingova funkcija definirana kot

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{M-1}\right). \quad (2.15)$$

Zatem so ti periodogrami, ki smo jih pomnožili s funkcijo okna, povprečeni v času

$$P_{xx}^w(f) = \frac{1}{L} \sum_{i=0}^{K-1} P_{xx}^i(f). \quad (2.16)$$

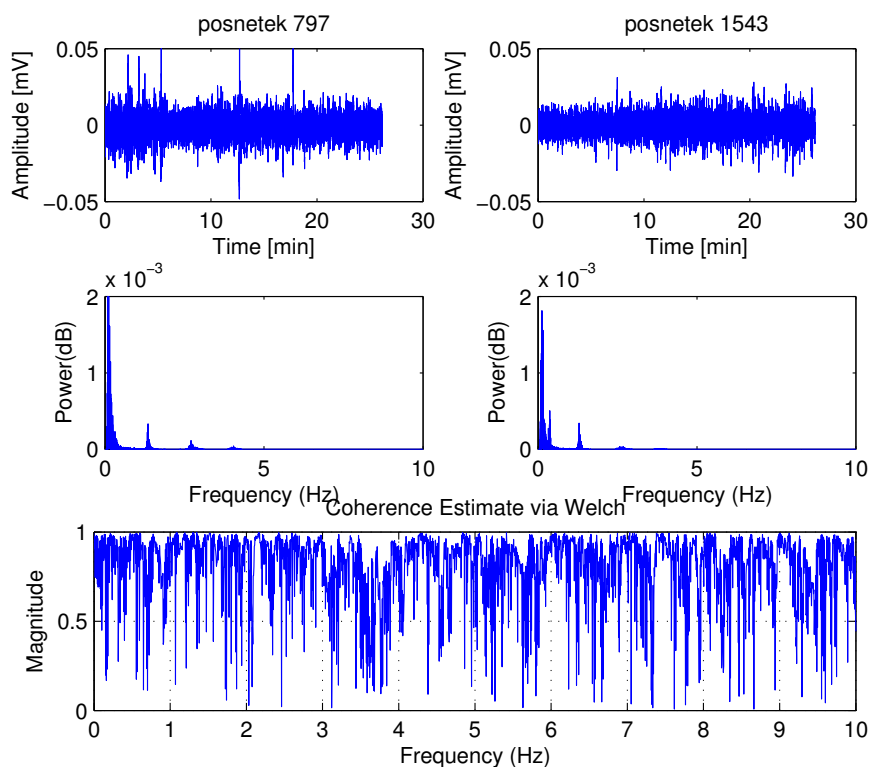
Da dobimo oceno navzkrižnega spektra, moramo povprečiti delne navzkrižne spektre

$$P_{xy}^i(f) = \frac{1}{E_w} \left( \sum_{n=0}^{M-1} x^i(n) w(n) e^{-2i\pi(f/f_{pr})n} \right) \cdot \left( \sum_{n=0}^{M-1} y^i(n) w(n) e^{2i\pi(f/f_{pr})n} \right). \quad (2.17)$$

V praksi je za prekrivanje signalov pogosto izbrana vrednost  $D = M/2$ , medtem ko je število intervalov  $K$  izbrano kot kompromis med resolucijo spektra in natančnostjo lociranja.

### 2.4.2 Vrednotenje koherenčne funkcije

Izračun koherenčne funkcije na dveh različnih signalih iz baze TPEHG DB nam ne da veliko uporabnih informacij. Signali maternične aktivnosti so časovno popolnoma neodvisni en od drugega, zato smo računali koherenčno funkcijo na močnostnih spektrih signalov. Močnostni spekter signala je moč signala pri vsaki frekvenci. Na ta način nam izračunana koherenca dejansko pove nekaj o tem, koliko je skupnih frekvenc in posledično, koliko sta si ta dva signala med seboj podobna.



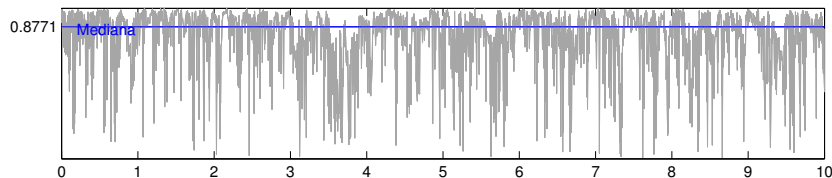
Slika 2.6: 1. signala, 2. močnostna spektra 3. koherenčna funkcija

Če primerjamo koherenčni funkciji na slikah 2.5 in 2.6 opazimo, da je koherentnost večja pri izračunu koherence močnostnih spektrov signalov kot samo signalov. Prav tako vidimo, da sta si grafa močnostnih spektrov na sliki 2.6 mnogo bolj podobna kot grafa osnovnih signalov.

Koherenčna funkcija nam za vsako izbrano frekvenco vrne oceno, koliko sta si dva signala med seboj podobna - koherentna. Vrednost te ocene je med 0 in 1. Za nadaljno analizo in pridobivanje rezultatov pa smo potrebovali eno vrednost koherenčne funkcije za celotno frekvenčno območje izbranih signalov. Za cenilki smo izbrali mediano amplitude koherence in integral koherence močnostnih spektrov signalov.

Mediana je numerična vrednost, ki razdeli podatke, populacijo ali verjetnostno porazdelitev na večjo in manjšo polovico. Pri računanju mediane amplitude koherence smo imeli končni nabor vrednosti koherence, za vsako frekvenco v danem frekvenčnem območju. Mediano potem izračunamo z urejanjem numeričnih vrednosti po velikosti in za mediano izberemo vrednost na polovici. Za nas je vse to opravila vgrajena MATLAB funkcija, ki vrne vrednost mediane M, za dan seznam A:

`M = median(A)`



Slika 2.7: Mediana amplitude koherence močnostnih spektrov posnetkov 747 in 1543

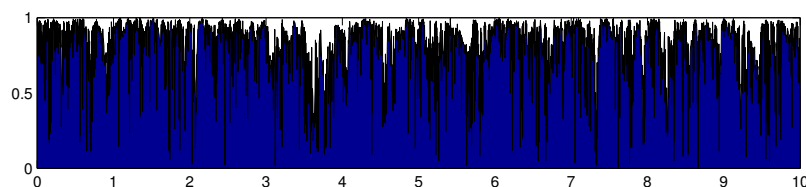
Integral koherenčne funkcije smo izbrali za drugo cenilko. Predstavlja površino območja omejenega z x-osjo, grafom koherenčne funkcije  $f(x)$  ter navpičnima premicama  $x=a$  in  $x=b$ .

$$\int_a^b f(x)dx$$



Torej bolj sta si dva signala podobna, večja bo koherenca in zatorej večja vrednost integrala. Seveda je vrednost integrala odvisna od mejnih vrednosti frekvenčnega območja  $a$  in  $b$ . Ti dve vrednosti variirata zaradi uporabe filtrov z različnimi mejnimi frekvencami. Primerjave med različnimi skupinami posnetkov smo zato delali samo pri signalih filtriranih z istimi mejnimi frekvencami. Za izračun integrala funkcije  $X$ , smo uporabili trapezoidno numerično integracijo, ki je v MATLAB integrirana z naslednjo funkcijo:

```
I = trapz(X)
```



Slika 2.8: Integral koherence močnostnih spektrov posnetkov 747 in 1543

## 2.5 Analiza variance ANOVA

Pri analizi variance se ukvarjamo z ugotavljanjem, ali vzorci izhajajo iz ene populacije ali iz več različnih populacij glede na ocenjene vrednosti variance. Opisne spremenljivke v tem primeru so kategorijske spremenljivke, ki jih imenujemo tudi faktorji. Glede na stopnje znotraj faktorjev in glede na število faktorjev poznamo različne postopke analize variance. Uporabili smo enosmerno analizo variance.

Enosmerna analiza variance temelji na primerjavi aritmetičnih sredin in nam pove, ali se aritmetične sredine vzorčnih skupin med sabo statistično pomembno razlikujejo. Če se vzorčne aritmetične sredine med sabo statistično pomembno razlikujejo, lahko to posplošimo na osnovno množico. Med katerimi podskupinami prihaja do statistično pomembnih razlik,

nam povedo post-hoc testi [19].

Predpostavke ANOVE:

- normalna porazdelitev odvisnih spremenljivk na vseh populacijah,
- enakost varianc,
- vzorci so neodvisni in naključno izbrani.

Ničelna hipoteza pri ANOVI vedno predpostavlja, da so srednje vrednosti vseh populacij enake,  $H_0 : \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_n$ . Hipotezo zavrnilo v primeru, če obstaja populacija z različno srednjo vrednostjo. Temu pravimo nasprotna hipoteza  $H_a : \exists \mu_i, \mu_j : \mu_i \neq \mu_j$ .

Testiranje hipoteze pri ANOVI poteka podobno kot pri ostalih statističnih testih in ga imenujemo F-razmerje in ga označimo s F. F-razmerje računamo pri številu populacij (v nadaljevanju skupin)  $a$  in  $a > 2$ .

$$F = \frac{MS_b}{MS_w} \quad (2.18)$$

$MS_b$ : Srednja vrednost med skupinami (ang. mean square between)

$$MS_b = \frac{SS_b}{f_b} \quad (2.19)$$

$SS_b$ : Vsota kvadratov med skupinami (ang. sum of square between)

$$SS_b = \sum (\bar{Y}_i - \bar{Y})^2 \quad (2.20)$$

$\bar{Y}_i$ : Srednja vrednost posamezne skupine

$\bar{Y}$ : Srednja vrednost vseh skupin

$f_b$ : Prostostna stopnja med skupinami

$$f_b = a - 1 \quad (2.21)$$

$MS_w$ : Srednja vrednost znotraj skupine (ang. mean square within)

$$MS_w = \frac{SS_w}{f_w} \quad (2.22)$$

$SS_w$ : Vsota kvadratov znotraj skupine (ang. sum of square within)

$$SS_w = \sum_i^a \sum_j^n (Y_j - \bar{Y}_i)^2 \quad (2.23)$$

$f_w$ : Prostostna stopnja znotraj skupine

$$f_w = a(n - 1) \quad (2.24)$$

Za določanje kritične meje F distribucije si pomagamo s spodnjo tabelo, kjer za vrednosti  $f_b$  in  $f_w$  poiščemo kritično mejo. Če je vrednost F manjša od vrednosti iz tabele, potem lahko ničelno tezo zavržemo in ugotovimo, da se skupine med seboj pomembno razlikujejo. Dodatno lahko izračunamo tudi p-vrednost, to je verjetnost, da ob predpostavki ničelne hipoteze dobimo vrednost F.

## 2.6 Klasifikacija posnetkov

Cilj raziskav, ki se ukvarjajo z napovednim modelom za prezgodnji porod, je čim boljša natančnost, čim bolj zgodaj v nosečnosti. S tem omogočimo, da zajamemo večji delež nosečnic, pri katerih lahko pride do zapletov, dovolj zgodaj, da imamo le-te možnost preprečiti.

Naloga klasifikatorja je za objekt, opisan z množico značilk, določiti, kateremu izmed možnih razredov pripada. Značilke so neodvisne zvezne ali diskretne spremenljivke, s katerimi opisujemo objekte, razred pa je odvisna diskretna spremenljivka, ki ji določimo vrednost glede na vrednosti neodvisnih spremenljivk. Zato da lahko klasifikator določi razred, mora imeti na nek način predstavljeno diskretno funkcijo, ki preslika prostor atributov v razred. Ta funkcija je lahko podana vnaprej ali pa kot v našem primeru naučena iz podatkov. Naloga učnega algoritma je torej iz množice vzorcev in z njim znanimi pripadajočimi razredi zgraditi pravilo, ki ga bo klasifikator uporabil pri klasifikaciji prihodnjih primerov [17].

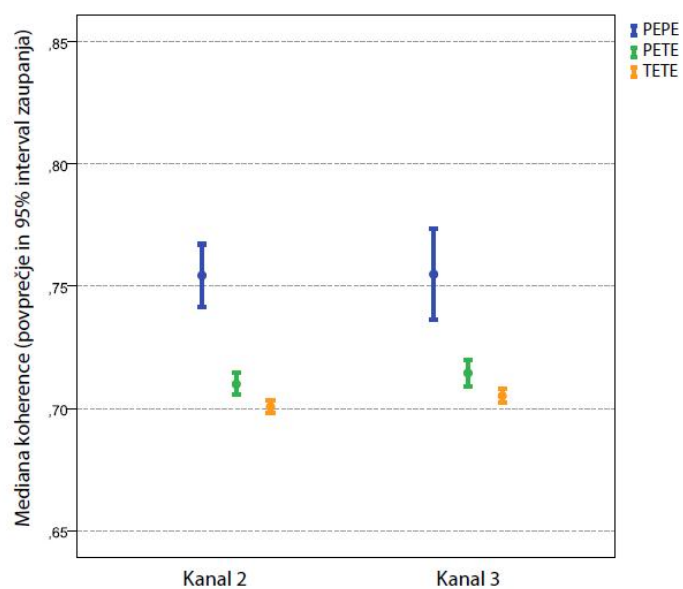
Za preizkušanje cenilk mediane amplitude in integrala koherence močnostnega spektra, smo uporabili Bayesov klasifikator, lastno razvit klasifikator in odločitvena drevesa. Za testiranje klasifikatorjev smo uporabili že vgrajena orodja iz MATLABA, ki smo jih preuredili glede na svoje potrebe. Naivni Bayesov klasifikator smo izbrali zaradi preproste implementacije in nizke računske zahtevnosti, odločitvena drevesa pa zaradi dobrih rezultatov klasifikacije pri podobnih raziskavah.

Cenilki, ki smo jih uporabili pri klasifikaciji, sta med seboj povezani. Zaradi medsebojne povezanosti pričakujemo tudi podobne klasifikacijske rezultate.

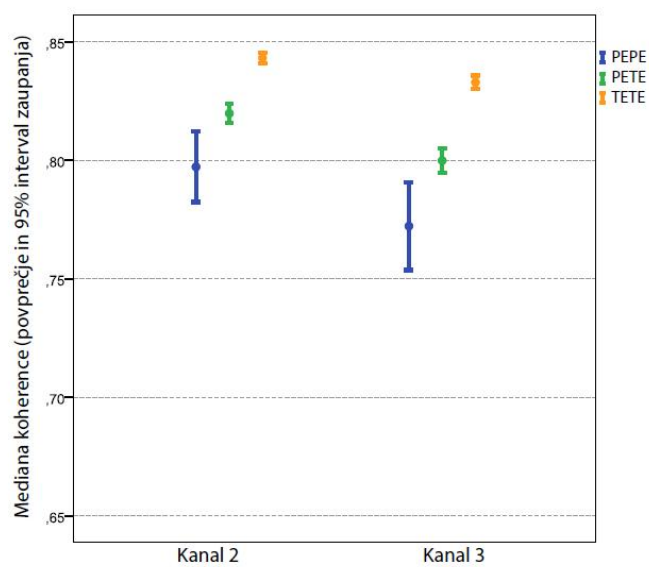
### **2.6.1 Klasifikator z uporabo značilk v različnih frekvenčnih območjih**

Sestavili smo ga empirično na podlagi rezultatov, ki smo jih pridobili z analizo ANOVE. Klasificirali smo izbrane signale v frekvenčnih območjih, ki so pokazala statistično pomembno ločevanje med skupinami prezgodnjega in terminskega poroda. Pri statistični analizi smo ugotovili, da se skupine PP, TT in PT med seboj razlikujejo na dva načina:

1. Srednja vrednost mediane koherence je najnižja znotraj skupine prezgodnjih porodov (PP), sledi srednja vrednost koherence med terminskim in prezgodnjim porodom (PT) in največja je znotraj terminskih porodov (TT). Slika: 2.9
2. Srednja vrednost mediane koherence je najnižja znotraj skupine terminskih porodov (TT), sledi srednja vrednost koherence med terminskim in prezgodnjim porodom (PT) in največja je znotraj prezgodnjih porodov (PP). Slika: 2.10



Slika 2.9: Koherenca je najvišja med posnetki prezgodnjih porodov.



Slika 2.10: Koherenca je najvišja med posnetki terminskih porodov.

Pri primerjavi vrednosti koherence med dvema izbranimi filtroma smo opazili, da se vrednosti za eno skupino močno razlikujejo, medtem ko se za drugo minimalno spremenijo. To ugotovitev smo uporabili pri gradnji lastnega klasifikatorja.

Klasificiramo neznani posnetek  $X$ . Za  $X$  imamo izračunani povprečni vrednosti koherence s skupino prezgodnjih porodov ( $P$ ) in s skupino terminskih porodov ( $T$ ) za oba izbrana frekvenčna območja. V spodnji kodi, ti dve vrednosti podamo v parametrih  $p1$ ,  $t1$  za prvo območje in  $p2$ ,  $t2$  za drugo območje. Vsak od parametrov je matrika s 3 stolpci. Prvi predstavlja številko posnetka, drugi predstavlja povprečno vrednost koherence, izračunano z vsemi iz skupine prezgodnjih porodov in tretji stolpec je povprečna vrednost koherence z vsemi iz skupine terminskih porodov.

Učenje klasifikatorja poteka tako, da izračunamo razliko koherence med frekvenčnima območjema za skupino prezgodnjih posnetkov in za skupino s terminskimi posnetki. Pri vsaki od skupin moramo izračunati dve razliki, tisto znotraj skupine ( $dpp$  in  $dt$ ) in razliko z drugo skupino ( $dpt$  in  $dt$ ). Za obe razliki posamezne skupine izračunamo srednjo vrednost in dobimo povprečno vrednost spremembe vrednosti koherence med frekvenčnima območjema. Mejo za klasifikacijo postavimo med obe srednji vrednosti posameznih skupin. Ker je sprememba vrednosti koherence, pri skupini s prezgodnjimi porodi manjša od skupine s terminskimi porodi, klasificiramo posnetek kot prezgodnji porod, če je izračunana razlika pod določeno mejo. V obratnem primeru ga klasificiramo kot posnetek s terminskim porodom. Glede na dejansko in napovedano skupino posnetka, napoved pade v enega izmed 4 razredov metrike zmogljivosti  $M$ .

```
function [M] = klasifikator(p1,t1,p2,t2)

%Matrika zmogljivosti
M = zeros(2,2);

%Ucenje klasifikatorja
dpp = (p1(:,2)-p2(:,2));
dpt = (p1(:,3)-p2(:,3));

dtp = (t1(:,2)-t2(:,2));
dtt = (t1(:,3)-t2(:,2));

prezgodnji = mean(dpp+dpt);
terminski = mean(dtp+dtt);

meja = (prezgodnji + terminski)/2;

%klasifikacija 19 posnetkov s prezgodnjim porodom
for i=1:19
    x = (p1(i,2)-p2(i,2))+(p1(i,3)-p2(i,3));
    if(x<meja)
        M(1,1) = M(1,1)+1;
    else
        M(1,2) = M(1,2)+1;
    end
end

%klasifikacija 142 posnetkov s terminskim porodom
for i=1:142
    x = (t1(i,2)-t2(i,2))+(t1(i,3)-t2(i,3));
    if(x<meja)
        M(2,1) = M(2,1)+1;
    else
        M(2,2) = M(2,2)+1;
    end
end
```

### 2.6.2 Naivni Bayesov klasifikator

Naivni Bayesov klasifikator temelji na predpostavki, da so atributi primera, ki ga klasificiramo, neodvisni med seboj. Če ima vsak primer v učni množici lastnosti  $F_{i,...,k}$  in pripada enemu iz množice med seboj tujih razredov  $C$ , lahko za vse primere izračunamo verjetnosti, da pripadajo danim razredom. S pomočjo Bayesovega izreka lahko z enačbo 2.25 pri danih atributih neznanega primera izračunamo verjetnost, da pripada razredu  $C$  [4].

$$p(C|F) = \frac{p(F_1|C)p(F_2|C)...p(F_k|C)p(C)}{p(F)} \quad (2.25)$$

Pri tem smo privzeli, da so atributi  $F_{i,...,k}$  med seboj neodvisni.

Ker je imenovalec  $p(F)$  vnaprej neznan, ga moramo iz izraza izločiti. Na srečo imenovalec ni odvisen od razreda  $C$ . To pomeni, da ga lahko iz različnih ocen verjetnosti pripadanja posameznim razredom izpostavimo. Recimo, da ima ciljna spremenljivka  $m$  možnih vrednosti  $c_1, c_2, ..., c_m$  [4]. Naivni Bayesov klasifikator za vsak razred  $c_j$  po naivni Bayesovi formuli izračuna verjetnost, da primer z naborom lastnosti  $F_{i,...,k}$  pripada razredu  $c_j$ . Primer klasificira v razred z največjo verjetnostjo [17]. Če velja 2.26 sklepamo, da dani primer z naborom lastnosti  $F_{i,...,k}$  pripada razredu  $c$ :

$$p(c) = \max\left(\prod_{i=1}^k p(f_i|c_j)\right); c_j \in C \quad (2.26)$$

### 2.6.3 Odločitvena drevesa

Odločitvena drevesa štejemo v skupino metod nadzorovanega avtomatskega učenja na podlagi rešenih primerov. Učenje deluje po metodi deli in vlada, atributi pa se izbirajo s pomočjo hevristične funkcije. Uporaba hevristik zmanjša računsko zahtevnost gradnje optimalnega odločitvenega drevesa. Tehnika odločitvenih dreves je dobila ime po diagramu, ki je prikazan v obliki drevesa. Diagram je sestavljen iz treh bistvenih komponent - vozlišč, vej in listov [18]. V vozliščih drevesa se nahajajo pogoji, v



katerih nastopajo dejstva, v listih pa možne odločitve, kaj storiti. Vsaka odločitev v poljubnem listu drevesa predstavlja izbiro, da dani primer pripada nekemu razredu [4].

Postopek za gradnjo odločitvenega drevesa je po [14, 18] naslednji:

1. Če so vsi objekti v  $C$  člani istega razreda, ali če je  $C$  prazna množica, označimo vozlišče kot list, mu določimo razred ter končamo postopek. V primeru, da je  $C$  prazna množica, v list namesto razreda zapišemo 'OdločitevNiMogoča'.
2. Naj bo  $A_i$  atribut izbran s pomočjo heuristične funkcije v trenutnem vozlišču z vrednostmi  $A_{i1}, A_{i2}, \dots, A_{in}$ . Vozlišče označimo kot notranje in vanj ustavimo atribut  $A_i$ .
3. S pomočjo vrednosti atributa razdelimo  $C$  na podmnožice  $C_{i1}, C_{i2}, \dots, C_{in}$ . Vsaka podmnožica  $C$  vsebuje tiste objekte iz  $C$ , ki imajo vrednost atributa  $A_i$  enako  $A_{in}$ .
4. Za vsako vrednost atributa in ustrezno podmnožico  $C$  ustvarimo vozlišče.
5. Za vsako ustvarjeno vozlišče ponovimo celoten postopek gradnje s  $C$  in z množico še ne uporabljenih atributov na poti do ustvarjenega vozlišča.

#### 2.6.4 Ocena delovanja klasifikatorja

Klasifikacija je potekala med posnetki, ki so bili snemani zgodaj in med posnetki, ki so bili snemani pozno v nosečnosti. Klasificirali smo jih v dva razreda - na tiste, kjer je prišlo do prezgodnjega poroda in razred posnetkov, kjer je bil porod ob roku.

Za oceno delovanja klasifikatorja smo uporabili klasično matriko zmogljivosti, iz katere smo izračunali občutljivost, specifičnost in natančnost klasifikatorja. Pri klasični matriki zmogljivosti dogodek, ki ga napovedujemo označimo kot pozitiven in drugi je negativen. Prezgodnji porod

smo označili, kot pozitiven dogodek in terminski porod, kot negativen. Za vsakega od dogodkov klasifikator uvrsti dogodek v pravi razred ali v napačni razred. Na ta način so možne 4 skupine napovedi [20]:

- Pozitiven dogodek klasificiran v pravi razred, pripada skupini TP (ang. true positive).
- Pozitiven dogodek klasificiran v napačni razred, pripada skupini FN (ang. false negative).
- Negativen dogodek klasificiran v pravi razred, pripada skupini TN (ang. true negative).
- Negativen dogodek klasificiran v napačni razred, pripada skupini FP (ang. false positive).

		Napovedan razred	
		Pozitiven	Negativen
Pravi razred	Pozitiven	TP	FN
	Negativen	FP	TN

Tabela 2.1: Klasična matrika zmogljivosti

Kombinacije možnih izidov so v tabeli 2.1. Vrednosti klasične matrike zmogljivosti, so osnova za izračun pogosto uporabljenih mer zmogljivosti klasifikatorja, ki so [20]:

### Občutljivost (ang. Sensitivity, Se)

Razmerje med pravilno napovedanimi pozitivnimi dogodki in vsemi pozitivno napovedanimi dogodki. V našem primeru je to razmerje vseh pravilno napovedanih prezgodnjih porodov proti vsem napovedanim prezgodnjim porodom.

$$Se = \frac{TP}{TP + FN} \quad (2.27)$$

**Specifičnost (ang. Specificity, Sp)**

Razmerje med pravilno napovedanimi negativnimi dogodki in vsemi negativnimi dogodki. V našem primeru je to razmerje vseh pravilno napovedanih terminskih porodov proti vsem napovedanim terminskim porodom.

$$Sp = \frac{TN}{TN + FP} \quad (2.28)$$

**Natančnost (ang. accuracy, Acc)**

Delež vseh pravilno napovedanih dogodkov med vsemi dogodki.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.29)$$

Kako dobro bo Bayesov klasifikator deloval, je odvisno tudi od tega, kako velika je učna množica in kateri primeri bodo sestavljali testno množico. Vse to pa je pogojeno velikosti baze primerov za klasifikacijski problem, ki jo imamo na voljo. Tako imamo različne načine sestavljanja učne množice, ki poizkušajo zaobiti težave, ki jih povzroča predvsem ne dovolj velika količina primerov.

V naši bazi posnetkov TPEHG DB sta najbolj problematični skupini posnetkov prezgodnjih porodov. Obe skupini snemano pozno, kot snemano zgodaj sestavlja le 19 posnetkov in to je v primerjavi s skupinama s terminskim porodom, kjer imamo na voljo 142 oz. 118 posnetkov, znatno manj. Zaradi manjšega števila posnetkov prezgodnjih porodov se klasifikator ne nauči dovolj dobro in posledično sledi slabša klasifikacija. Zaradi narave problema, kjer je delež prezgodnjih porodov le 6,2 % ne moremo pričakovati, da bi imeli popolno bazo. To problematiko poizkušamo zaobiti z različnimi načini klasifikacije.

**Ocena delovanja po principu učna - testna množica**

Ta način se uporablja, kadar imamo veliko število dostopnih primerov in kjer so skupine enako močne. Vse primere naključno razdelimo na dve skupini. Iz prve sestavimo učno množico, na kateri naučimo

klasifikator, ki potem klasificira drugo skupino - testno množico. Kljub temu da ta način ni najbolj primeren za našo bazo, je vseeno dobra referenca za primerjave z ostalimi načini .

### **Ocena delovanja na učni množici**

Kadar nimamo možnosti, da bi delovanje klasifikatorja preverili na neznanih primerih ali če želimo hitro oceno, lahko preverimo, ali klasifikator v prave razrede razvrsti primere, na katerih smo ga učili. To pomeni, da sta učna in testna množica enaki. Običajno je delovanje klasifikatorjev na učni množici boljše kot na neznanih primerih. Uspešnost klasifikatorja na učni množici torej lahko štejemo za nekakšno zgornjo mejo uspešnosti klasifikatorja [4].

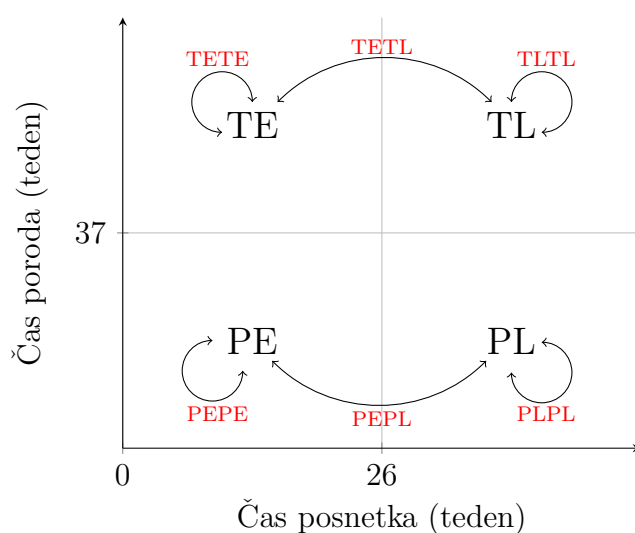
### **Ocena delovanja na način "izpusti enega"**

Ta pristop uporabimo, če je število dostopnih primerov majhno, tako kot v našem primeru. Tako za vsak primer  $i$  iz množice vseh primerov  $X$  klasifikator naučimo na množici  $X \setminus i$  in primer  $i$  klasificiramo s pomočjo naučenega klasifikatorja. S tem načinom omogočimo klasifikatorju dovolj veliko učno množico, hkrati pa za klasifikacijo lahko uporabimo vse dostopne primere.

# Rezultati

Preden smo lahko podatke statistično obdelali za pridobitev rezultatov, smo najprej morali izračunati cenilke za vse možne kombinacije parov signalov za vsak filter in za vsak kanal posnetka. Izračun koherence zahteva enako dolžino med signaloma. Zaradi krajše dolžine signala smo posnetka št. 1202 (TL) in 873 (TE) zavrgli. Uporabljenih je bilo tako 298 posnetkov, za vsak posnetek imamo 3 kanale in 9 filtrov različnih pasovnih dolžin. Pare signalov smo generirali iz naslednjih 6 parov skupin:

- PE - PE
- PL - PL
- PE - TE
- PL - TL
- TE - TE
- TL - TL



Vse skupaj smo tako dobili  $2 \cdot \binom{19}{2} + 19 \cdot 142 + \binom{142}{2} + 19 \cdot 118 + \binom{118}{2} = 22196$  različnih parov posnetkov. Če upoštevamo še 9 filtrov in 3 kanale, smo na koncu pridobili  $22196 \cdot 3 \cdot 9 = 599292$  različnih vrednosti cenilk mediane koherence in integrala.

pos. 1	skupina	pos. 2	skupina	filter	kanal	mediana	integral
1027	TL	734	TL	0,3-2,5	S3	0,92671	1811,32
642	TL	734	TL	0,3-2,5	S1	0,92466	1743,02
642	TL	734	TL	0,3-2,5	S2	0,9904	1981,69
642	TL	734	TL	0,3-2,5	S3	0,9678	1905,42
1526	PE	1600	PE	2,5-4,5	S1	0,82443	1559,81
1526	PE	1600	PE	2,5-4,5	S2	0,70735	1317,43
1526	PE	1600	PE	2,5-4,5	S3	0,59612	1137,47
1526	PE	786	PE	2,5-4,5	S1	0,86643	1642,82

Tabela 3.1: Izsek iz tabele podatkov

### 3.1 Ločevanje skupin

Na podlagi izračunanih srednjih vrednosti mediane amplitude koherence in integrala koherence ter pripadajočih standardnih deviacij, smo z uporabo Anove izračunali ali se izbrane skupine med seboj pomembno razlikujejo. V spodnjih tabelah so izračunane p-vrednosti med skupinami

- PEPE - PETE
- PLPL - PLTL
- TETE - PETE
- TLTL - PLTL

tako za integral kot za mediano amplitude. Ločujemo skupino posnetkov s terminskim porodom od skupine posnetkov s prezgodnjim porodom. Predpostavimo, da imamo posnetek X, za katerega ne vemo, kateri skupini pripada. Izračunamo koherenco s posnetkom terminskega (T) in koherenco s posnetkom prezgodnjega poroda (P). Dobimo dve skupini - XP in XT, za kateri si želimo, da sta statistično različni. Posnetek X

pripada enemu izmed razredov P ali T. Cilj je, da so statistično čimbolj različne naslednje skupine posnetkov: PP - PT in TP - TT.

Vrednosti, pri katerih smo dosegli oba pogoja, so v spodnjih tabelah odebeljene. Dobili smo frekvenčna območja in signal, pri katerem naj bi uporaba koherence za ločevanje posnetkov dala najboljše rezultate.

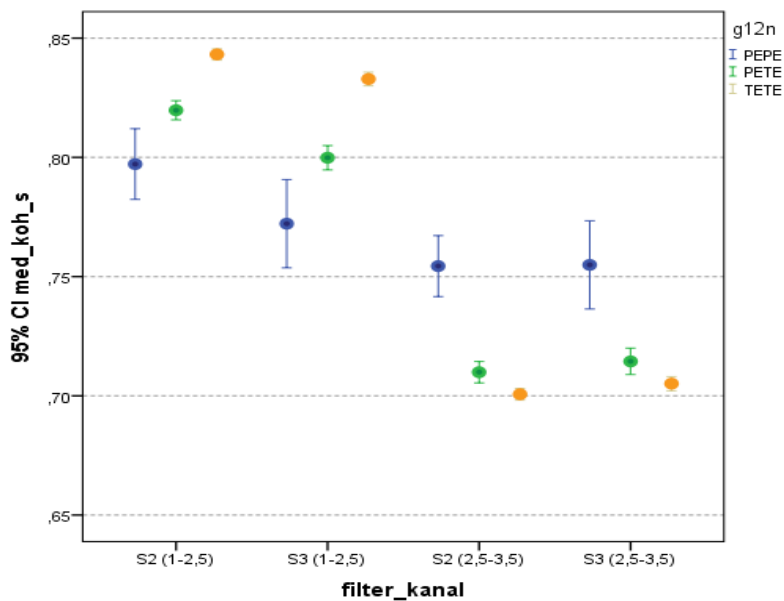
Statistično razlikovanje skupin smo opazovali pri štirih različnih skupinah. Skupine se razlikujejo po metodi ocenjevanja koherence - mediana amplitude ali integral ter po času snemanja signala - snemano zgodaj ali snemano pozno v nosečnosti.

### 3.1.1 Ločevanje zgodaj snemanih posnetkov z uporabo mediane amplitude

	PEPE – PETE			PETE – TETE		
Filter\Signal	S1	S2	S3	S1	S2	S3
0,08 – 4	0,448	1,0	0,136	0,0	1,0	0,0
0,3 – 3	1,0	1,0	0,272	0,813	1,0	0,0
0,3 – 4	1,0	1,0	0,425	0,1	0,004	0,0
0,3 – 1	0,349	0,321	0,601	0,0	0,0	0,0
1 – 2,5	0,315	<b>0,044</b>	<b>0,044</b>	0,0	<b>0,0</b>	<b>0,0</b>
2,5 – 3,5	0,004	<b>0,0</b>	<b>0,002</b>	0,110	<b>0,002</b>	<b>0,013</b>
3,5 – 4,5	0,469	1,0	1,0	0,0	0,174	0,035
0,3 – 2,5	0,849	1,0	1,0	0,003	0,134	0,488
2,5 – 4,5	1,0	1,0	1,0	0,029	0,086	0,302

Tabela 3.2: p-vrednosti za statistično ločevanje med skupinami glede na mediano amplitude koherence posnetkov snemanih zgodaj

Pasovni območji, kjer se je pokazala ločljivost med skupinami sta filtrirani s filtrom na prepustnimi frekvencami od 1 do 2,5 Hz in od 2,5 do 3,5 Hz. Pri obeh filterih se je pokazala boljša ločljivost pri kanalu signala S2.



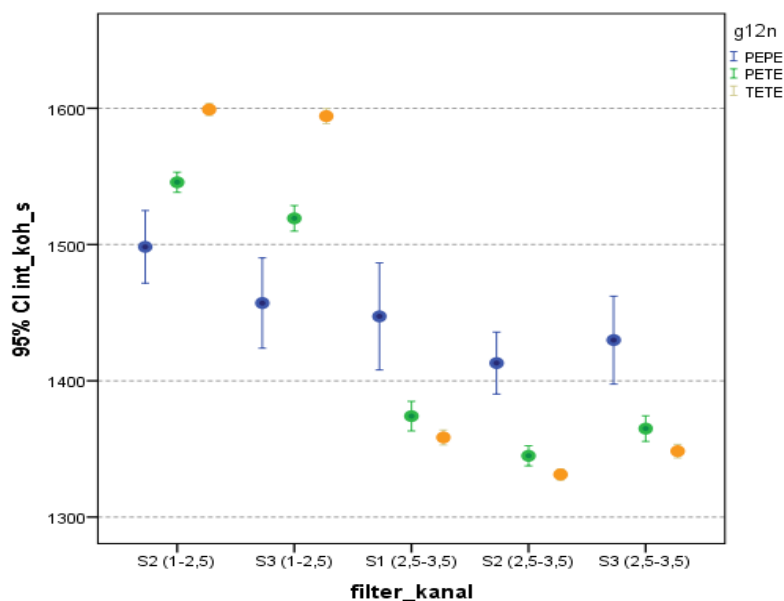
Slika 3.1: Intervali zaupanja za izbrana frekvenčna območja in signale

### 3.1.2 Ločevanje zgodaj snemanih posnetkov z uporabo integrala

Filter\kanal	PEPE – PETE			PETE – TETE		
	S1	S2	S3	S1	S2	S3
0,08 – 4	0,589	1,0	0,077	0,0	0,459	0,0
0,3 – 3	1,0	1,0	0,089	1,0	0,852	0,0
0,3 – 4	1,0	0,69	0,332	0,033	0,0	0,0
0,3 – 1	0,06	0,222	0,111	0,0	0,0	0,0
1 – 2,5	0,418	<b>0,018</b>	<b>0,012</b>	0,0	<b>0,0</b>	<b>0,0</b>
2,5 – 3,5	<b>0,003</b>	<b>0,0</b>	<b>0,003</b>	<b>0,031</b>	<b>0,006</b>	<b>0,008</b>
3,5 – 4,5	0,796	1,0	1,0	0,028	1,0	0,314
0,3 – 2,5	0,212	0,585	1,0	0,0	0,0	0,128
2,5 – 4,5	1,0	1,0	1,0	0,012	0,141	0,464

Tabela 3.3: p-vrednosti za statistično ločevanje med skupinami za integral koherence posnetkov snemanih zgodaj





Slika 3.2: Intervali zaupanja za izbrane filtre in kanale

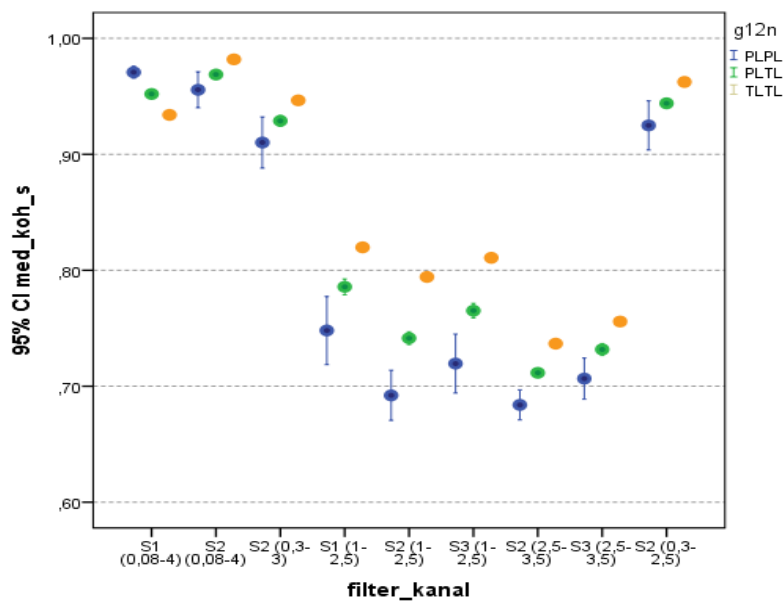
Uporaba integrala za vrednotenje koherence nam da podobne rezultate kot mediana. Statistične razlike med skupinami so se kot pri mediani amplitude pokazali pri frekvenčnih območjih 1-2,5 Hz in 2,5-3,5 Hz. Dodatno je bil za ločevanje skupin primeren tudi tretji signal, pri frekvenčnem območju 2,5-3,5 Hz.

### 3.1.3 Ločevanje pozno snemanih posnetkov z uporabo mediane amplitude

Pri ločevanju posnetkov snemanih pozno, smo ugotovili, da se skupine med seboj bolj razlikujejo. Za ločevanje je primernih pet frekvenčnih območij. To je bilo pričakovano zaradi tega, ker so signali časovno bližje poroda. Ponovno sta se izkazali frekvenčni območji, ki sta bili dobri za ločevanje skupin pri posnetkih snemanih zgodaj. Dodatna frekvenčna območja, ki so pokazala razlike med skupina pa so 0,3-2,5 Hz, 0,08-4 Hz in 0,3-3 Hz. Od signalov se je najbolj izkazal signal S2.

Filter\kanal	PLPL – PLTL			PLTL – TLTL		
	S1	S2	S3	S1	S2	S3
0,08 – 4	<b>0,046</b>	<b>0,001</b>	0,117	<b>0,0</b>	<b>0,0</b>	0,0
0,3 – 3	0,580	<b>0,006</b>	1,0	0,001	<b>0,0</b>	0,843
0,3 – 4	0,548	0,475	1,0	0,0	0,0	1,0
0,3 – 1	0,741	0,457	0,637	0,001	0,001	0,0
1 – 2,5	<b>0,004</b>	<b>0,0</b>	<b>0,0</b>	<b>0,0</b>	<b>0,0</b>	<b>0,0</b>
2,5 – 3,5	0,1	<b>0,001</b>	<b>0,045</b>	0,0	<b>0,0</b>	<b>0,0</b>
3,5 – 4,5	1,0	1,0	1,0	1,0	0,0	0,12
0,3 – 2,5	1,0	<b>0,001</b>	1,0	0,082	<b>0,0</b>	0,035
2,5 – 4,5	1,0	1,0	1,0	0,368	1,0	1,0

Tabela 3.4: p-vrednosti za statistično ločevanje med skupinami glede na mediano koherence posnetkov snemanih pozno



Slika 3.3: Intervali zaupanja za izbrane filtre in kanale

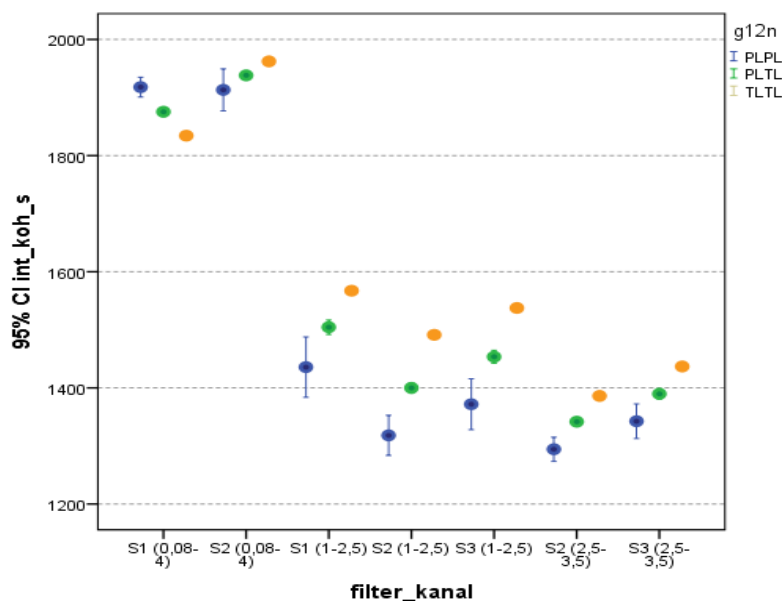
### 3.1.4 Ločevanje pozno snemanih posnetkov z uporabo integrala

Uporaba integrala je dala slabše rezultate kot mediana amplitude pri pozno snemanih posnetkih. Za ločevanje so statistično primerna tri frekvenčna območja. Tudi tukaj sta dobre rezultate dosegla območji 1-2,5Hz in 2,5-3,5Hz in pa še območje 0,08-4Hz. Znova je bil najboljši signal S2.

V splošnem se je integral kot metoda za vrednotenje koherenčne funkcije izkazal malo slabše. Zaradi podobnega načina ocenjevanja koherence, so vsa frekvenčna območja, ki so se pokazala kot primerna za ločevanje skupin pri uporabi integrala, zajeta tudi pri mediani amplitude.

	PLPL – PLTL			PLTL - TLTL		
Filter\kanal	S1	S2	S3	S1	S2	S3
0,08 – 4	<b>0,044</b>	<b>0,046</b>	0,232	<b>0,0</b>	<b>0,0</b>	0,0
0,3 – 3	0,486	0,083	1,0	0,0	0,0	1,0
0,3 – 4	0,280	0,951	1,0	0,0	0,015	0,374
0,3 – 1	0,415	1,0	0,476	0,0	0,005	0,0
1 – 2,5	<b>0,005</b>	<b>0,0</b>	<b>0,0</b>	<b>0,0</b>	<b>0,0</b>	<b>0,0</b>
2,5 – 3,5	0,055	<b>0,001</b>	<b>0,029</b>	0,0	<b>0,0</b>	<b>0,0</b>
3,5 – 4,5	1,0	1,0	1,0	1,0	0,0	0,012
0,3 – 2,5	0,849	0,031	1,0	0,009	0,0	0,013
2,5 – 4,5	1,0	1,0	1,0	0,133	1,0	1,0

Tabela 3.5: p-vrednosti za statistično ločevanje med skupinami glede na integral koherence posnetkov snemanih pozno



Slika 3.4: Intervali zaupanja za izbrane filtre in kanale

## 3.2 Ocenjevanje klasifikacije

Posnetke smo klasificirali z uporabo koherenčne funkcije. Posnetek smo klasificirali v dve skupini - skupino posnetkov, kjer se nosečnost konča s prezgodnjim porodom ali v skupino, kjer se nosečnost konča s porodom ob roku. Klasifikacija je potekala na treh različnih klasifikatorjih. Najboljšo klasifikacijo smo dosegli z odločitvenim drevesom, za katerega je značilno učinkovito prilagajanje učni množici. Klasifikator, ki smo ga razvili sami, je dosegel nekoliko slabšo oceno klasifikacije, vendar pa so ti rezultati bolj realni, glede na zahtevnost problema klasificiranja. Klasifikacija z Bayesovim klasifikatorjem ni bila uspešna. Večino posnetkov je klasificiral v skupino terminskih porodov.

Način izbire učne in testne množice je imel velik vpliv na oceno klasifikacije. Ocena delovanja klasifikatorja na učni množici je bila boljša od ostalih dveh načinov. V realnosti poizkušamo klasificirati neznane

primere, zato lahko ta način štejemo za zgornjo mejo uspešnosti. Bolj realna pokazatelja uspešnosti klasifikatorja pa sta načina po principu izpusti enega oziroma učna - testna množica.

### 3.2.1 Ocena zmogljivosti delovanja klasifikatorja z uporabo značilk v različnih frekvenčnih območjih

SIGNAL	ČAS	METODA	NAČIN	Se	Sp	Acc
S2	E	mediana	učna	0.63	0.65	0.65
S3	E	mediana	učna	0.58	0.61	0.61
S2	E	integral	učna	0.63	0.63	0.63
S3	E	integral	učna	0.58	0.63	0.63
S2	E	mediana	učna-testna	0.67	0.65	0.65
S3	E	mediana	učna-testna	0.56	0.62	0.61
S2	E	integral	učna-testna	0.67	0.63	0.64
S3	E	integral	učna-testna	0.56	0.63	0.63
S2	E	mediana	izpusti enega	0.63	0.65	0.65
S3	E	mediana	izpusti enega	0.58	0.61	0.61
S2	E	integral	izpusti enega	0.63	0.63	0.63
S3	E	integral	izpusti enega	0.53	0.63	0.62

Tabela 3.6: Občutljivost in specifičnost lastnega klasifikatorja med frekvenčnima območjema 1-2,5 Hz in 2,5-3,5 Hz

### 3.2.2 Ocene zmogljivosti delovanja Bayesovega klasifikatorja

FILTER	ČAS SNEMANJA	METODA	SIGNAL	Se	Sp	Acc
1-2,5	E	Mediana	S2	0	0.88	0.86
1-2,5	E	Mediana	S3	0	0.88	0.85
2,5-3,5	E	Mediana	S2	0	0.89	0.88
2,5-3,5	E	Mediana	S3	0	0.89	0.88
0,08-4	L	Mediana	S1	0.14	0.88	0.49
0,08-4	L	Mediana	S2	0.09	0.86	0.74
0,3-2,5	L	Mediana	S2	0	0.86	0.79
0,3-3	L	Mediana	S2	0	0.86	0.82
1-2,5	L	Mediana	S1	0.25	0.88	0.84
1-2,5	L	Mediana	S2	0.33	0.89	0.84
1-2,5	L	Mediana	S3	0.4	0.88	0.84
2,5-3,5	L	Mediana	S2	0.25	0.86	0.83
2,5-2,3	L	Mediana	S3	0	0.86	0.84
1-2,5	E	Integral	S2	0	0.88	0.86
1-2,5	E	Integral	S3	0	0.88	0.85
2,5-3,5	E	Integral	S1	0	0.9	0.89
2,5-3,5	E	Integral	S2	0	0.9	0.89
2,5-3,5	E	Integral	S3	0	0.9	0.89
0,08-4	L	Integral	S1	0.13	0.87	0.69
0,08-4	L	Integral	S2	0.09	0.86	0.74
1-2,5	L	Integral	S1	0.25	0.88	0.84
1-2,5	L	Integral	S2	0.33	0.89	0.84
1-2,5	L	Integral	S3	0.25	0.88	0.84
2,5-3,5	L	Integral	S2	0.25	0.88	0.84
2,5-3,5	L	Integral	S3	0	0.87	0.85

Tabela 3.7: Občutljivost, specifičnost in natančnost Bayesovega klasifikatorja po principu učna - testna množica

FILTER	ČAS SNEMANJA	METODA	SIGNAL	Se	Sp	Acc
1-2,5	E	Mediana	S2	0	1	0.88
1-2,5	E	Mediana	S3	0	1	0.88
2,5-3,5	E	Mediana	S2	0	1	0.88
2,5-3,5	E	Mediana	S3	0	1	0.88
0,08-4	L	Mediana	S1	0.79	0.25	0.32
0,08-4	L	Mediana	S2	0.05	0.95	0.82
0,3-2,5	L	Mediana	S2	0.05	0.92	0.8
0,3-3	L	Mediana	S2	0.05	0.95	0.82
1-2,5	L	Mediana	S1	0.16	0.93	0.82
1-2,5	L	Mediana	S2	0.26	0.94	0.85
1-2,5	L	Mediana	S3	0.21	0.96	0.85
2,5-3,5	L	Mediana	S2	0	1	0.86
2,5-2,3	L	Mediana	S3	0	1	0.86
1-2,5	E	Integral	S2	0	1	0.88
1-2,5	E	Integral	S3	0	1	0.88
2,5-3,5	E	Integral	S1	0	1	0.88
2,5-3,5	E	Integral	S2	0	1	0.88
2,5-3,5	E	Integral	S3	0	1	0.88
0,08-4	L	Integral	S1	0	1	0.86
0,08-4	L	Integral	S2	0.05	0.93	0.81
1-2,5	L	Integral	S1	0.16	0.93	0.82
1-2,5	L	Integral	S2	0.26	0.93	0.84
1-2,5	L	Integral	S3	0.21	0.96	0.85
2,5-3,5	L	Integral	S2	0	1	0.86
2,5-3,5	L	Integral	S3	0	1	0.86

Tabela 3.8: Občutljivost, specifičnost in natančnost Bayesovega klasifikatorja na učni množici

FILTER	ČAS SNEMANJA	METODA	SIGNAL	Se	Sp	Acc
1-2,5	E	Mediana	S2	0	0.99	0.88
1-2,5	E	Mediana	S3	0	0.97	0.86
2,5-3,5	E	Mediana	S2	0	1	0.88
2,5-3,5	E	Mediana	S3	0	1	0.88
0,08-4	L	Mediana	S1	0.79	0.25	0.32
0,08-4	L	Mediana	S2	0	0.95	0.82
0,3-2,5	L	Mediana	S2	0	0.92	0.79
0,3-3	L	Mediana	S2	0	0.94	0.81
1-2,5	L	Mediana	S1	0.16	0.93	0.82
1-2,5	L	Mediana	S2	0.26	0.93	0.84
1-2,5	L	Mediana	S3	0.21	0.96	0.85
2,5-3,5	L	Mediana	S2	0	1	0.86
2,5-2,3	L	Mediana	S3	0	1	0.86
1-2,5	E	Integral	S2	0	0.99	0.88
1-2,5	E	Integral	S3	0	0.97	0.86
2,5-3,5	E	Integral	S1	0	1	0.88
2,5-3,5	E	Integral	S2	0	1	0.88
2,5-3,5	E	Integral	S3	0	1	0.88
0,08-4	L	Integral	S1	0	1	0.86
0,08-4	L	Integral	S2	0	0.93	0.8
1-2,5	L	Integral	S1	0.16	0.93	0.82
1-2,5	L	Integral	S2	0.21	0.92	0.82
1-2,5	L	Integral	S3	0.21	0.96	0.85
2,5-3,5	L	Integral	S2	0	1	0.86
2,5-3,5	L	Integral	S3	0	1	0.86

Tabela 3.9: Občutljivost, specifičnost in natančnost Bayesovega klasifikatorja po principu "izpusti enega"



**3.2.3 Ocene zmogljivosti delovanja odločitvenega drevesa**

FILTER	ČAS SNEMANJA	METODA	SIGNAL	Se	Sp	Acc
1-2,5	E	Mediana	S2	0.11	0.9	0.81
1-2,5	E	Mediana	S3	0.11	0.9	0.81
2,5-3,5	E	Mediana	S2	0.11	0.9	0.81
2,5-3,5	E	Mediana	S3	0.11	0.9	0.81
0,08-4	L	Mediana	S1	0.22	0.9	0.81
0,08-4	L	Mediana	S2	0.44	0.93	0.87
0,3-2,5	L	Mediana	S2	0.22	0.92	0.82
0,3-3	L	Mediana	S2	0.22	0.92	0.82
1-2,5	L	Mediana	S1	0.22	0.92	0.82
1-2,5	L	Mediana	S2	0.22	0.9	0.81
1-2,5	L	Mediana	S3	0.22	0.9	0.81
2,5-3,5	L	Mediana	S2	0.11	0.9	0.79
2,5-2,3	L	Mediana	S3	0.11	0.88	0.78
1-2,5	E	Integral	S2	0.11	0.9	0.81
1-2,5	E	Integral	S3	0.11	0.9	0.81
2,5-3,5	E	Integral	S1	0.11	0.9	0.81
2,5-3,5	E	Integral	S2	0.11	0.89	0.8
2,5-3,5	E	Integral	S3	0.11	0.89	0.8
0,08-4	L	Integral	S1	0.22	0.92	0.82
0,08-4	L	Integral	S2	0.33	0.93	0.85
1-2,5	L	Integral	S1	0.11	0.88	0.78
1-2,5	L	Integral	S2	0.22	0.92	0.82
1-2,5	L	Integral	S3	0.11	0.9	0.79
2,5-3,5	L	Integral	S2	0.22	0.92	0.82
2,5-3,5	L	Integral	S3	0.22	0.88	0.79

Tabela 3.10: Občutljivost, specifičnost in natančnost odločitvenega drevesa po principu učna - testna množica

FILTER	ČAS SNEMANJA	METODA	SIGNAL	Se	Sp	Acc
1-2,5	E	Mediana	S2	0.68	0.98	0.94
1-2,5	E	Mediana	S3	0.74	0.99	0.96
2,5-3,5	E	Mediana	S2	0.68	0.98	0.94
2,5-3,5	E	Mediana	S3	0.74	0.99	0.96
0,08-4	L	Mediana	S1	0.84	0.98	0.96
0,08-4	L	Mediana	S2	0.89	0.97	0.96
0,3-2,5	L	Mediana	S2	0.95	0.98	0.98
0,3-3	L	Mediana	S2	0.89	0.98	0.97
1-2,5	L	Mediana	S1	0.68	0.96	0.92
1-2,5	L	Mediana	S2	0.68	0.98	0.94
1-2,5	L	Mediana	S3	0.89	0.91	0.91
2,5-3,5	L	Mediana	S2	0.68	0.97	0.93
2,5-2,3	L	Mediana	S3	0.79	0.98	0.96
1-2,5	E	Integral	S2	0.53	0.99	0.94
1-2,5	E	Integral	S3	0.53	0.99	0.93
2,5-3,5	E	Integral	S1	0.68	0.99	0.96
2,5-3,5	E	Integral	S2	0.74	0.96	0.94
2,5-3,5	E	Integral	S3	0.63	0.98	0.94
0,08-4	L	Integral	S1	0.89	0.97	0.96
0,08-4	L	Integral	S2	0.84	0.97	0.96
1-2,5	L	Integral	S1	0.74	0.96	0.93
1-2,5	L	Integral	S2	0.79	0.97	0.94
1-2,5	L	Integral	S3	0.95	0.95	0.95
2,5-3,5	L	Integral	S2	0.53	0.98	0.92
2,5-3,5	L	Integral	S3	0.79	0.99	0.96

Tabela 3.11: Občutljivost, specifičnost in natančnost odločitvenega drevesa na učni množici

FILTER	ČAS SNEMANJA	METODA	SIGNAL	Se	Sp	Acc
1-2,5	E	Mediana	S2	0.16	0.91	0.82
1-2,5	E	Mediana	S3	0.26	0.9	0.83
2,5-3,5	E	Mediana	S2	0.21	0.9	0.82
2,5-3,5	E	Mediana	S3	0.05	0.9	0.8
0,08-4	L	Mediana	S1	0.32	0.88	0.8
0,08-4	L	Mediana	S2	0.42	0.92	0.85
0,3-2,5	L	Mediana	S2	0.42	0.95	0.88
0,3-3	L	Mediana	S2	0.42	0.93	0.86
1-2,5	L	Mediana	S1	0.21	0.94	0.84
1-2,5	L	Mediana	S2	0.26	0.95	0.85
1-2,5	L	Mediana	S3	0.16	0.86	0.76
2,5-3,5	L	Mediana	S2	0.16	0.95	0.84
2,5-2,3	L	Mediana	S3	0.11	0.93	0.82
1-2,5	E	Integral	S2	0.05	0.94	0.83
1-2,5	E	Integral	S3	0.11	0.91	0.81
2,5-3,5	E	Integral	S1	0.05	0.94	0.83
2,5-3,5	E	Integral	S2	0.05	0.88	0.78
2,5-3,5	E	Integral	S3	0	0.85	0.75
0,08-4	L	Integral	S1	0.21	0.93	0.83
0,08-4	L	Integral	S2	0.47	0.92	0.86
1-2,5	L	Integral	S1	0.16	0.92	0.81
1-2,5	L	Integral	S2	0.21	0.95	0.85
1-2,5	L	Integral	S3	0	0.91	0.78
2,5-3,5	L	Integral	S2	0.05	0.94	0.82
2,5-3,5	L	Integral	S3	0.37	0.94	0.86

Tabela 3.12: Občutljivost, specifičnost in natančnost odločitvenega drevesa po principu "izpusti enega"

## Zaključek

Osnovni opis koherenčne funkcije - poišče skupne frekvence dveh signalov in jih ovrednoti, je ustrezal metodi, ki bi učinkovito ločila posnetke prezgodnjega in terminskega poroda. Zaradi kompleksnosti EMG signalov maternice smo frekvenčna območja razdelili na manjša podobmočja. S tem smo omogočili, da morebitne frekvence, ki nosijo informacije o času poroda, pridejo bolj do izraza in s tem izboljšajo ločevanje posnetkov prezgodnjega in terminskega poroda. Pri ločevanju skupin prezgodnjega in terminskega poroda, snemanih zgodaj, sta najbolj pokazljivi frekvenčni območji 1-2,5 Hz in 2,5-3,5 Hz. Rezultati so zanimivi, ker je do sedaj veljalo, da se večina dogajanja povezanega s prezgodnjim porodom dogaja pri nizkih frekvencah do 1 Hz. Druga izredno zanimiva ugotovitev je obrat vrednosti, tako mediane amplitude, kot integrala koherence med skupinama. Do obrata pride izključno pri zgodaj snemanih posnetkih. Vrednost koherence je največja pri frekvenčnem območju od 1-2,5 Hz znotraj skupine terminskih porodov, pri frekvenčnem območju od 2,5-3,5 Hz pa je največja znotraj skupine prezgodnjih porodov. To lahko nakazuje, da frekvence med 2,5-3,5 Hz vsebujejo informacije, ki bi bile značilne za prezgodnji porod. To je odlična iztočnica za že opravljene preiskave, ki so se omejevale na širša frekvenčna področja. Pri posnetkih snemanih pozno, je več frekvenčnih območij, ki so pokazljiva za ločevanje skupin. To smo pričakovali, ker so posnetki bližje porodu in tako vsebujejo že več aktivnosti maternice povezane s porodom. Tudi pri pozno snemanih, sta pokazljiva oba območja 1-2,5 Hz in 2,5-3,5 Hz, poleg teh pa še območja 0,08-4Hz, 0,3-3Hz in 0,3-2,5 Hz. Pri skupini posnetkov snemanih pozno,

obrata vrednosti koherence nismo zaznali.

Obe cenilki koherentne funkcije, mediana amplitude in integral, sta pričakovano pokazali podobne rezultate. Uporaba integrala je pri posnetkih snemanih zgodaj pokazala da je za ločevanje skupin primeren tudi tretji signal na frekvenčnem območju 2,5-3,5Hz. Pri posnetkih snemanih pozno se je malo bolje izkazala metoda mediane amplitude, ki je pokazala, da so za ločevanje primerni tudi območji 0,3-3 Hz in 0,3-2,5 Hz drugega signala. Tu je ostalo še veliko prostora, za testiranje ostalih cenilk koherentne funkcije.

Najboljšo oceno delovanja Bayesovega klasifikatorja smo dosegli pri posnetkih snemanih pozno, na frekvenčnem območju od 1-2,5 Hz pri drugem signalu, tako pri mediani amplitude kot pri integralu. Občutljivost klasifikacije je bila med 26 % in 33 %, specifičnost med 89 % in 93 % in natančnost med 82 % in 85 % za vse tri načine klasifikacije. V splošnem se klasifikacija z uporabo Bayesovega klasifikatorja ni izkazala zaradi občutljivosti, ki je blizu 0 % in nasprotno specifičnost blizu 100 %. To pomeni, da smo vse posnetke klasificirali kot terminske, kar je nasprotno od naših želja po ločevanju prezgodnjega od terminskega poroda.

Odločitvena drevesa imajo dobro sposobnost prilagajanja in zato so rezultati klasifikacije boljši kot pri Bayesovem klasifikatorju. Kako dobro se lahko odločitvena drevesa prilagodijo opazimo pri načinu klasifikacije 'učna množica', kjer smo dosegli 95 % občutljivost, 98 % specifičnost in 98 % natančnost pri drugem signalu na frekvenčnem območju 0,3-2,5 Hz. Tudi pri ostalih izbranih frekvenčnih območjih se občutljivost in specifičnost približata 100 %. Pri načinih 'učna-testna množica' in 'izpusti enega' dosežemo bolj realne rezultate z občutljivostjo do 44 % in specifičnostjo do 93 %. V splošnem je ocena klasifikacije boljša pri posnetkih snemanih pozno, kar je bil pričakovan rezultat.

Klasifikacija s klasifikatorjem z uporabo značilk v različnih frekvenčnih območjih je dala spodbudne rezultate. Za vse tri načine klasifikacije je bila občutljivost med 53 % in 63 %, občutljivost med 61 % in 65 % in

natančnost med 61 % in 65 %. To velja za pare frekvenčnih območij, pri katerih pride do obrata v vrednosti koherenčne funkcije.

Vsi trije klasifikatorji delujejo samo na osnovi koherenčne funkcije brez dodatnih kliničnih informacij, kot so tlak v maternici, starost ženske, število preteklih porodov, teža, hipertenzija, diabetes, kajenje, krvavenje v nosečnosti. Z upoštevanjem teh parametrov bi bilo mogoče klasificirati posnetke znatno bolje.

V prihodnosti bi bilo potrebno razširiti bazo TPEHG DB in s tem omogočiti testiranje naših hipotez in preteklih raziskav na novih primerih. Tako bi lahko z gotovostjo trdili, katere metode so dobre in katere so se pretirano prilagodile primerom iz baze.

# Dodatek

## A Statistični rezultati ANOVE

Tabela 5.1: Srednje vrednosti mediane koherenčne funkcije izračunane med posnetki znotraj skupine PE in pripadajoča standardna deviacija

Kanal Filter [Hz]	PE - PE					
	S1		S2		S3	
	SV	SD	SV	SD	SV	SD
0,08 - 4	0,92	0,07	0,97	0,04	0,95	0,06
0,3 - 3	0,70	0,14	0,84	0,12	0,78	0,15
0,3 - 4	0,86	0,07	0,93	0,05	0,91	0,07
0,3 - 1	1,0	0,0	1,0	0,0	1,0	0,0
1 - 2,5	0,80	0,17	0,80	0,10	0,77	0,12
2,5 - 3,5	0,76	0,14	0,75	0,08	0,75	0,12
3,5 - 4,5	0,66	0,24	0,69	0,17	0,68	0,20
0,3 - 2,5	0,90	0,06	0,95	0,04	0,94	0,06
2,5 - 4,5	0,51	0,18	0,46	0,20	0,46	0,20

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.3: Srednje vrednosti mediane koherenčne funkcije izračunane med posnetki znotraj skupine TE in pripadajoča standardna deviacija

Kanal	TE - TE					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	0,89	0,14	0,97	0,06	0,92	0,11
0,3 - 3	0,71	0,19	0,86	0,14	0,74	0,17
0,3 - 4	0,85	0,13	0,93	0,08	0,88	0,11
0,3 - 1	1,0	0,0	1,0	0,0	1,0	0,00
1 - 2,5	0,85	0,16	0,84	0,12	0,83	0,15
2,5 - 3,5	0,71	0,16	0,70	0,13	0,71	0,15
3,5 - 4,5	0,69	0,17	0,70	0,14	0,70	0,16
0,3 - 2,5	0,92	0,11	0,96	0,06	0,93	0,10
2,5 - 4,5	0,49	0,15	0,45	0,14	0,48	0,14

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.2: Srednje vrednosti mediane koherenčne funkcije izračunane med posnetki skupin PE in TE in pripadajoča standardna deviacija

Kanal	PE - TE					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	0,90	0,11	0,97	0,05	0,94	0,09
0,3 - 3	0,71	0,17	0,85	0,13	0,76	0,16
0,3 - 4	0,86	0,11	0,93	0,07	0,90	0,10
0,3 - 1	1,0	0,0	1,0	0,0	1,0	0,00
1 - 2,5	0,82	0,17	0,82	0,11	0,80	0,13
2,5 - 3,5	0,72	0,17	0,71	0,12	0,71	0,15
3,5 - 4,5	0,68	0,20	0,69	0,16	0,69	0,17
0,3 - 2,5	0,91	0,09	0,95	0,05	0,93	0,08
2,5 - 4,5	0,50	0,17	0,46	0,17	0,47	0,16

SV - srednja vrednost, SD - standardna deviacija



Tabela 5.5: Srednje vrednosti mediane koherenčne funkcije izračunane med posnetki skupin PL in TL in pripadajoča standardna deviacija

Kanal	PL - TL					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	0,95	0,08	0,97	0,08	0,97	0,05
0,3 - 3	0,80	0,18	0,89	0,14	0,82	0,16
0,3 - 4	0,90	0,11	0,93	0,11	0,92	0,09
0,3 - 1	1,0	0,0	1,0	0,0	1,0	0,00
1 - 2,5	0,79	0,17	0,74	0,13	0,77	0,15
2,5 - 3,5	0,72	0,15	0,71	0,09	0,73	0,12
3,5 - 4,5	0,64	0,21	0,64	0,14	0,65	0,19
0,3 - 2,5	0,92	0,10	0,94	0,10	0,93	0,08
2,5 - 4,5	0,53	0,17	0,51	0,15	0,52	0,17

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.4: Srednje vrednosti mediane koherenčne funkcije izračunane med posnetki znotraj skupine PL in pripadajoča standardna deviacija

Kanal	PL - PL					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	0,97	0,03	0,96	0,10	0,98	0,02
0,3 - 3	0,82	0,20	0,88	0,17	0,82	0,18
0,3 - 4	0,91	0,11	0,91	0,15	0,91	0,09
0,3 - 1	1,0	0,0	1,0	0,0	1,0	0,0
1 - 2,5	0,75	0,19	0,69	0,14	0,72	0,17
2,5 - 3,5	0,70	0,15	0,68	0,09	0,71	0,12
3,5 - 4,5	0,65	0,19	0,64	0,13	0,65	0,18
0,3 - 2,5	0,93	0,10	0,92	0,14	0,93	0,08
2,5 - 4,5	0,52	0,15	0,51	0,15	0,51	0,16

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.7: Srednje vrednosti integrala koherenčne funkcije izračunane med posnetki znotraj skupine PE in pripadajoča standardna deviacija

Kanal	PE - PE					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	1774	194	1909	138	1870	167
0,3 - 3	1323	256	1597	247	1484	282
0,3 - 4	1624	161	1794	140	1755	174
0,3 - 1	2048	0	2048	0	2048	0
1 - 2,5	1567	321	1498	177	1457	220
2,5 - 3,5	1447	260	1413	150	1430	214
3,5 - 4,5	1286	415	1325	302	1324	342
0,3 - 2,5	1716	157	1844	125	1822	157
2,5 - 4,5	1018	312	938	341	940	328

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.6: Srednje vrednosti mediane koherenčne funkcije izračunane med posnetki znotraj skupine TL in pripadajoča standardna deviacija

Kanal	TL - TL					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	0,93	0,10	0,98	0,03	0,96	0,06
0,3 - 3	0,79	0,16	0,90	0,11	0,82	0,13
0,3 - 4	0,89	0,12	0,95	0,06	0,92	0,09
0,3 - 1	1,0	0,0	1,0	0,0	1,0	0,00
1 - 2,5	0,82	0,14	0,79	0,11	0,81	0,13
2,5 - 3,5	0,76	0,15	0,74	0,10	0,76	0,13
3,5 - 4,5	0,64	0,23	0,66	0,14	0,66	0,19
0,3 - 2,5	0,91	0,10	0,96	0,04	0,94	0,08
2,5 - 4,5	0,53	0,19	0,51	0,16	0,52	0,17

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.9: Srednje vrednosti integrala koherenčne funkcije izračunane med posnetki znotraj skupine TE in pripadajoča standardna deviacija

Kanal	TE - TE					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	1719	274	1919	158	1790	239
0,3 - 3	1361	320	1647	278	1408	300
0,3 - 4	1628	253	1804	177	1679	226
0,3 - 1	2048	0	2048	0	2048	0
1 - 2,5	1649	306	1599	226	1594	281
2,5 - 3,5	1358	280	1331	208	1348	255
3,5 - 4,5	1330	296	1334	240	1342	269
0,3 - 2,5	1778	227	1873	150	1806	207
2,5 - 4,5	988	239	924	224	963	217

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.8: Srednje vrednosti integrala koherenčne funkcije izračunane med posnetki skupin PE in TE in pripadajoča standardna deviacija

Kanal	PE - TE					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	1747	241	1914	149	1829	210
0,3 - 3	1344	291	1623	264	1446	294
0,3 - 4	1627	211	1800	160	1717	205
0,3 - 1	2048	0	2048	0	2048	0
1 - 2,5	1603	318	1546	195	1519	249
2,5 - 3,5	1374	287	1345	196	1365	249
3,5 - 4,5	1313	351	1330	271	1332	304
0,3 - 2,5	1748	193	1859	138	1815	184
2,5 - 4,5	1004	275	934	279	956	267

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.11: Srednje vrednosti integrala koherenčne funkcije izračunane med posnetki skupin PL in TL in pripadajoča standardna deviacija

Kanal	PL - TL					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	1875	187	1938	186	1907	136
0,3 - 3	1535	331	1723	290	1568	291
0,3 - 4	1732	236	1811	233	1764	206
0,3 - 1	2048	0	2048	0	2048	0
1 - 2,5	1504	301	1400	218	1454	271
2,5 - 3,5	1383	259	1342	154	1390	214
3,5 - 4,5	1245	364	1223	224	1253	322
0,3 - 2,5	1789	226	1855	218	1817	196
2,5 - 4,5	1050	288	1023	253	1034	276

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.10: Srednje vrednosti integrala koherenčne funkcije izračunane med posnetki znotraj skupine PL in pripadajoča standardna deviacija

Kanal	PL - PL					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	1918	113	1913	239	1930	92
0,3 - 3	1576	365	1703	331	1575	329
0,3 - 4	1758	232	1780	289	1759	217
0,3 - 1	2048	0	2048	0	2048	0
1 - 2,5	1436	344	1318	229	1372	291
2,5 - 3,5	1334	256	1294	138	1343	198
3,5 - 4,5	1264	324	1224	202	1245	303
0,3 - 2,5	1808	229	1823	283	1803	209
2,5 - 4,5	1034	247	1013	244	1025	257

SV - srednja vrednost, SD - standardna deviacija

Tabela 5.12: Srednje vrednosti integrala koherenčne funkcije izračunane med posnetki znotraj skupine TL in pripadajoča standardna deviacija

Kanal	TL - TL					
	S1		S2		S3	
Filter	SV	SD	SV	SD	SV	SD
0,08 - 4	1834	231	1962	102	1886	166
0,3 - 3	1494	297	1740	240	1558	248
0,3 - 4	1708	235	1841	156	1768	191
0,3 - 1	2048	0	2048	0	2048	0
1 - 2,5	1567	263	1491	194	1538	239
2,5 - 3,5	1450	266	1386	169	1437	236
3,5 - 4,5	1254	393	1261	225	1276	327
0,3 - 2,5	1773	219	1884	123	1829	179
2,5 - 4,5	1065	321	1024	256	1033	288

SV - srednja vrednost, SD - standardna deviacija

# Literatura

- [1] I. Verdenik, Ž. Novak-Antolič, J. Zupan, *Perinatologia Slovenica II : slovenski perinatalni rezultati za obdobje 2002-11*. Združenje za perinatalno medicino SZD : Ginekološka klinika, UKC, 2013.
- [2] G. Fele-žorž, G. Kavšek, Ž. Novak-Antolič, F. Jager, *A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups*. Medical Biological Engineering Computing Sep, zv. 46, št. 9, str. 911-922, 2008
- [3] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, *Prediction of Preterm Deliveries from EHG Signals Using Machine Learning*. PLoS ONE 8(10): e77154. doi:10.1371/journal.pone.0077154, 2013
- [4] G. Fele-Žorž, *Nelinearna analiza signalov elektromiograma maternice za napovedovanje prezgodnjega poroda*. Doktorska disertacija, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, 2010
- [5] I. Verdenik, *Večplastni napovedni model prezgodnjega poroda*. Doktorska disertacija, Univerza v Ljubljani, Medicinska fakulteta, Ljubljana, 2002
- [6] H. Leman, C. Marque, J. Gondry, *Use of the electrohysterogram signal for characterization of contractions during pregnancy*. IEEE

- Transactions on Biomedical Engineering, 46(10):1222-1229, Clinical Trial, 1999
- [7] P. D. Welch, *The use of fast Fourier transform for estimation of power spectra: A method based on time averaging over short, modified periodograms*, *IEEE Transactions on Audio Electroacoustics*. AU-15, str. 70-73, 1967
- [8] JD. Iams *Prediction and early detection of preterm labor*. The American College of Obstetricians and Gynecologist, zv. 101, št. 2, str. 402-412, 2003
- [9] R. E. Garfield , W. L. Maner *Physiology and electrical activity of uterine contractions*. Semin Cell Dev Biology, št. 18, str. 289–295, 2007
- [10] A. Golińska *Coherence Function in Biomedical Signal Processing: a Short Review of Applications in Neurology, Cardiology and Gynecology*. Studies in Logic, Grammar and Rhetoric Second Special Issue on Logical, Statistical and Computer Methods in Medicine, zv. 25, št. 38, str. 73-81, 2011
- [11] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, *Prediction of Preterm Deliveries from EHG Signals Using Machine Learning*. PLoS ONE 8(10): e77154. doi:10.1371/journal.pone.0077154, 2013
- [12] J. Gondry, C. Marque, J. Duchene, D. Cabrol, *Electrohysterography during Pregnancy: Preliminary Report*. Biomedical instrumentation and technology association for the advancement of medical instrumentation, zv. 27, št. 4, str. 318–324, 1993
- [13] M. Lucovnik, W. L. Maner, L. R. Chambliss, R. Blumrick, J. Balducci, *Noninvasive uterine electromyography for prediction of*

- preterm delivery*. American journal of obstetrics and gynecology zv. 204, št. 3, str. 228.e1–10, 2011
- [14] M. Zorman, V. Podgorelec, M. Lenič, P. Povalej, P. Kokol, A. Tapa-jner *Inteligenti sistemi in profesionalni vsakdan*. Univerza v Mari-boru, Center za interdisciplinarne in multidisciplinarne raziskave in študije, str. 64, 2003
- [15] F. Jager, *Separating sets of term and pre-term uterine EMG records*. Predavanje na Massachusetts Institute of Technology, Cambridge, Massachusetts, 2013
- [16] J. G. Proakis, D. G. Manolakis, *Digital signal processing : principles, algorithms, and applications*. Upper Saddle River (New Jersey) : Prentice Hall, 1996
- [17] I. Čuček, *Filtriranje elektronske pošte z Bayesovim klasifikatorjem*. Diplomsko delo, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, 2010
- [18] S. Štraus, *Podatkovno rudarjenje na primeru zavarovalnice Maribor*. Diplomsko delo, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, 2012
- [19] <http://www.termania.net/slovarji/terminoloski-slovar-vzgoje-in-izobrazevanja/3474669/enosmerna-analiza-variance—anova>
- [20] J. P. Egan, *Signal Detection Theory and ROC-analysis*. Academic Press, Inc., 1975